ALLIANCE PARTICIPATION AND MILITARY SPENDING

A Dissertation

by

JOSHUA KEEGAN ALLEY

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Matthew Fuhrmann
Committee Members,	Hyeran Jo
	William R. Clark
	Jasen J. Castillo
Head of Department,	Guy D. Whitten

August 2020

Major Subject: Political Science

ABSTRACT

In this dissertation, I examine how participation in military alliances impacts military spending. Despite expectations that alliances affect military spending, we still do not know how. One group of scholars claims that alliance participation often decreases military spending, but another group expects that alliance participation increases military spending. I offer an answer to this longstanding puzzle by explaining when alliance participation increases or decreases military spending. First, I reassess a well known framework for understanding alliances — the public goods model of alliances. Then, I describe how treaty depth modifies the impact of alliance participation on non-major power military spending. Last, I show that states add depth to alliance treaties to reassure their partners while reducing exposure to entrapment, so democracies often form deep alliances.

Taken together, the argument and evidence in this project indicate that efforts to establish credible alliance commitments shape the bargaining leverage of states, which then determines how alliance participation affects military spending. To start, I examine the public goods model of alliances, and find that differences in economic weight do not lead to differences in military spending. As a result, my argument relies on bargaining between alliance members, and claims that how alliance participation affects military spending depends on state capability and alliance treaty design. Specifically, treaty depth modifies how non-major power military spending responds to alliance participation. Treaty depth refers to the extent of defense cooperation formalized in the alliance through promises like policy coordination and basing rights. When security-seeking non-major powers join deep alliances they usually decrease military spending, because these treaties are more credible. Joining shallow alliances often increases non-major power military spending due to credibility concerns and higher allied leverage. Finally, I show that deep alliances are the result of efforts to increase the credibility of alliance commitments while managing the risk of entrapment. I argue that because treaty depth has limited domestic audience costs but reassures allies, democracies often form deep alliances to increase the credibility of their alliances while managing exposure to entrapment.

To test this argument, I make two innovations in research design. First, I provide a general and reliable test of the public goods theory of alliances, which is missing from previous research. Then, I build a multilevel model to show how alliance characteristics like treaty depth modify the impact of alliance participation on military spending. Multilevel modeling addresses limitations of previous research designs by estimating the unique effect of individual alliances.

DEDICATION

To Mary Ann Baner, a great teacher and trailblazer.

ACKNOWLEDGMENTS

Science is a collective endeavor, and this dissertation is no exception. First, my thanks to an outstanding committee. Matthew Fuhrmann provided patient, thorough feedback and guidance in honing both the content and presentation of this project. Hyeran Jo offered perceptive comments that helped me consider issues in new ways. Bill Clark taught me a great deal about how to approach grad school and political science and offered useful scrutiny. Jasen Castillo offered helpful comments and pushed me to clarify the policy implications of the research.

The Department of Political Science at Texas A&M also deserves acknowledgement and thanks. Participants in several internal workshops provided useful feedback. My special thanks to Molly Berkemeier, Florian Hollenbach, Alejandro Medina, Austin Mitchell, Ben Peterson and Carlisle Rainey for productive conversations.

Social scientists outside of Texas A&M offered generous support for this project. Caroline Hartzell encouraged me to pursue graduate study in political science. I benefited from comments at the 2019 Texas Triangle IR Conference, the 2019 ISA Meeting, and the 2019 Peace Science Society (International) Annual Meeting. Benjamin Fordham, Paul Hensel, Ashley Leeds, Erik Lin-Greenberg, and Paul Poast provided especially useful criticisms and suggestions. Participants in the 2018-2019 Oskar Morgenstern Fellowship at the Mercatus Center also engaged with several extensions of the project, especially the public goods paper.

I owe a great deal to my family. My parents, John and Mary supported my lifelong interest in history, politics and economics, and taught me the value of hard work and persistence. My grandparents also provided encouragement, not least by asking when I would finish my dissertation. This dissertation is dedicated to my grandmother, Mary Ann Baner, who was an exceptional teacher and academic pioneer. My siblings John and Phebe also offered regular and often timely support. Sarah, my wife, entered the picture later in graduate school and provided support, love and encouragement to finish well.

Finally and most importantly: *Soli Deo Gloria*.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supported by a dissertation committee consisting of Professor Matthew Fuhrmann [advisor], Professor Hyeran Jo and Professor William Roberts Clark of the Department of Political Science and Professor Jasen Castillo of the Bush School of Government and Public Service.

All other work conducted for the dissertation was completed by the student independently.

Funding Sources

Graduate study was supported by a stipend from Texas A&M University, and a research grant from the Department of Political Science.

NOMENCLATURE

NATO	The North Atlantic Treaty Organization

TABLE OF CONTENTS

Page

LIST OF FIGURES

FIGURE Page

LIST OF TABLES

TABLE Page

1. INTRODUCTION

How does alliance participation affect military spending? Scholars have long expected that joining international treaties where states promise support in war affects defense expenditures. Because domestic military capability and alliances both provide security, alliances and military spending could replace or complement each other.

Despite durable expectations that domestic arms and international alliances are related, we still do not know how alliance participation affects military spending due to an ongoing debate. Scholars agree that alliances impact military spending, but disagree about how. One side of the debate expects that alliance participation will reduce military spending. Another perspective asserts that alliance participation will increase military spending. These competing theoretical claims are buttressed by corresponding empirical findings. Therefore, alliance participation and military spending is a salient puzzle in alliance politics. It is unclear whether alliance participation increases or decreases military expenditures.

In addition to the academic puzzle, alliance participation and military spending is a political issue. Allied defense spending is a salient concern in US foreign policy. Complaints that US allies spend too little on their military are a hardy perennial across presidential administrations (Lanoszka, 2015). Barack Obama and Donald Trump have both criticized allied defense efforts in their own ways.

Understanding the costs and benefits of alliances also informs debates over the future of US foreign policy. Advocates of deep engagement (Brooks, Ikenberry and Wohlforth, 2013) and restraint (Posen, 2014) in grand strategy have different views of alliances. Proponents of restraint argue that the United States should withdraw from many alliances, because allies spend too little on defense, which necessitates higher US defense spending (Preble, 2009). Deep engagement advocates argue that the benefits of alliances exceed the costs and believe that the problem of low allied military spending is overstated (Brands and Feaver, 2017).

In this dissertation, I address the question of alliance participation and military spending. I

claim that alliance participation can increase or decrease military spending, depending on state capability and alliance treaty design. First, I show that the public goods model of alliances is unlikely to provide a satisfactory explanation of when alliance participation increases or decreases military spending. Then I use exchange and bargaining between alliance members to show that deep alliances decrease non-major power military spending, while participation in shallow alliances increases defense spending. Last I explain why states form deep alliances. I argue that states use depth to reassure allies while managing audience costs and entrapment, which means that democracies are more likely to employ treaty depth.

I conclude that how alliance participation affects military spending depends on exchange and bargaining between alliance members. Bargaining leverage for alliance members depends on how states establish credible commitments when the alliance forms. When alliance leaders use treaty depth to increase the credibility of the alliance, their junior partners have more freedom to reduce defense spending. As a result, alliance formation shapes bargaining leverage during alliance maintenance (Snyder, 1997).

I base these conclusions on a progression of arguments and research designs that addresses three related problems in previous scholarship on alliance participation and military spending. The first problem is largely theoretical. The second issue is mostly empirical. The third problem combines theory and empirics.

The first problem is that theories of alliance participation and military spending rarely acknowledge that alliances could increase or decrease military spending. In the only conditional argument, DiGiuseppe and Poast (2016) claim that democracies make more credible commitments, so defense pacts with democracies will be more likely to reduce military spending. Otherwise, scholars have not incorporated well-established findings that alliance treaty design and membership shape the consequences of alliance participation. For example, alliances have heterogeneous effects on conflict (Leeds, 2003*b*; Benson, 2012) and trade (Long, 2003; Long and Leeds, 2006). The same is likely true of military spending— it is possible that alliances could increase or decrease military spending, depending on alliance treaty design and membership.

Research design shortcomings are the second problem. Existing scholarship employs two types of research designs, both of which have important limitations. The first set of research designs compares states with at least one alliance to states without any alliances in a global sample e.g. (Conybeare, 1994; Goldsmith, 2003; Morgan and Palmer, 2006). These general research designs give broad estimates, but their state-level alliance measures mask heterogeneous effects of alliances. On the other hand, specific research designs estimate how changes in allied spending affect military spending in a few states e.g. (Barnett and Levy, 1991; Sorokin, 1994; Plümper and Neumayer, 2015). Specific studies capture the impact of participation in a important alliances, but it is hard to draw general conclusions from a study of one or two alliances.

The third issue is parallel theoretical and empirical development of public goods and bargaining models of alliances. These two strands of inquiry have progressed separately with distinct arguments and empirical strategies. Treating security from alliances as a public good is hard to reconcile with weighing the sources of bargaining leverage between members, as the public goods model claims small states have more leverage. The public goods (Olson and Zeckhauser, 1966) and joint product (Sandler and Hartley, 2001) models argue that security from an alliance is completely or partially a public good, which allows smaller alliance members to free-ride. These models of alliances use differences in state size, technology and doctrine to explain differences in military spending across alliance members. By contrast, bargaining models of alliance politics derive their explanatory power from political relations between members and institutional design. Bargaining arguments view alliances as institutions where states balance heterogeneous goals (Morrow, 1991; Snyder, 1997), negotiate joint war plans (Poast, 2019*a*) and have different sources of leverage. Public goods and bargaining approaches are hard to reconcile. In a public goods approach, context shifts the consequences of alliance participation for military spending. In exchange and bargaining, the relationship depends on which states participate in the alliance, what they use the treaty for, and how they design the alliance.

The relative parsimony of the public goods model gives it an outsized impact on scholarly and public debate, which belies an empirical concern. Despite the prominence of the public goods

model, there is little reliable and general empirical evidence of free-riding. Most of the public goods literature relies on misspecified correlations between GDP and defense burdens. The defense burden measures military spending as a share of GDP. Therefore, correlations between GDP and defense burdens place GDP on both sides of the equation, which biases inferences. As a result of this specification problem, there is little reliable evidence testing the well-known predictions of free-riding in alliances from Olson and Zeckhauser (1966)'s public goods model. A few studies address the specification problem (Plümper and Neumayer, 2015; George and Sandler, 2017), but these studies examine NATO, which is a difficult alliance to draw general conclusions from. The emphasis on defense burdens in tests of the public goods model also hinders comparisons with bargaining models. The effect of relative economic weight on defense burdens is hard to compare with the impact of changes in allied capability or alliance participation dummies on defense spending.

The three papers in this dissertation address the above theoretical and empirical issues. Theoretically, I develop an argument that uses differences in state capability and alliance treaty design to predict how alliance participation affects military spending. I also explore how bargaining between prospective alliance members produces different alliance treaty designs.

Empirically, I address shortcomings in previous research designs. To start, I implement a general and more reliable test of the public goods argument, and find little evidence of free-riding based on economic size. I then use a multilevel model to model the heterogeneous effects of individual alliances on military spending using alliance treaty design and membership. Multilevel modeling combines the strengths of specific and general research designs by estimating how states respond to changes in allied capability as a function of alliance treaty design and membership.

1.1 Plan of the Dissertation

In the following three chapters, I establish how alliance participation affects military spending. In Chapter 2, I assess the public goods perspective on alliances and military spending. First, I show that much existing work on the public goods model relies on misspecified correlations between GDP and defense burdens, which has hindered the accumulation of reliable and general evidence.

I employ the public goods logic to predict that states with a low share of GDP in an alliance will often decrease military spending, while states with a high share of allied GDP will often increase military spending. Then, I employ a variant of the multilevel model from Chapter 2 to test these predictions from the public goods model. I find little evidence that differences in economic weight drive changes in military spending. Based on this finding I conclude that theories of bargaining and exchange and between alliance members are better suited to explain heterogeneity in how alliance participation affects military spending.

Chapter 3 details the puzzle of alliance participation and military spending, and explains how treaty depth modifies the relationship between alliance participation and military spending among non-major powers. This chapter focuses on non-major powers to ensure theoretical and empirical parsimony. I argue that credibility from a treaty depth shapes whether alliance participation increases or decreases non-major power military spending. Deep alliances will decrease non-major power military spending, while participation in shallow alliances where allies have more leverage is more likely to increase military spending. The research design uses a Bayesian multilevel model to estimate how treaty membership and design lead to heterogeneous effects of alliance participation on military spending. I find that deep alliances often decrease non-major power military spending, but shallow alliances often increase defense expenditures. Based on this, I conclude that states trade off between reassurance and higher allied military spending in alliance politics.

Last, Chapter 4 explains when states form alliances with deep treaties. This chapter addresses concerns about endogeneity from non-random selection into alliances by showing that depth is the result of efforts to establish credible commitments. This paper also contributes to a nascent literature on alliance treaty design (Mattes, 2012*b*; Benson, 2012; Benson and Clinton, 2016) I argue that states use treaty depth to increase the credibility of their alliances while managing exposure to audience costs. This combination of greater credibility and lower audience costs is especially appealing to democracies, because democracies have higher audience costs from violating promises of military support. Therefore, I expect that as the democracy score of the leading alliance member rises, treaty depth will increase, but the probability of unconditional military support will decrease.

I find mixed evidence for these claims, as the association between democracy and unconditional military support is weaker than expected. I conclude that treaty depth is driven by credibility concerns, rather than efforts to manage allied military spending. Furthermore, because democracy encourages deep alliances, alliances with democracies are not fully limited commitments.

2. REASSESSING THE PUBLIC GOODS THEORY OF ALLIANCES

Olson and Zeckhauser (1966) argue that international alliances generate a collective action problem. According to their theory, security from an alliance is a public good, so smaller alliance participants "free ride" on the contributions of larger members. Free-riding is reflected by disproportionate allocations of resources to defense, where smaller alliance members spend a lower share of their national income on the military, relative to larger partners.

In this chapter, I test the free-riding prediction with a statistical model that addresses model specification and generalizability issues in previous research. I use the public goods logic to predict that states with a low share of total GDP in an alliance will decrease percentage changes in military spending, and states with high share of total allied GDP will increase military spending. Then I employ a Bayesian model to estimate the association between economic weight and percentage changes in military spending for 204 alliances. I find little evidence of free-riding based on economic weight.

There are three reasons to undertake another test of the public goods model. First, although academic theory has progressed since 1966, the public goods model retains an important place in discourse about alliances. The public goods model and related modifications have a salient place in scholarly debate e.g.,(Walt, 1990; Sandler, 1993; Mearsheimer, 1994; Goldstein, 1995; Sandler and Hartley, 2001; Garfinkel, 2004; Walt, 2009; Norrlof, 2010; Barrett, 2010; Plümper and Neumayer, 2015).

Second, policy and popular discussions of alliances employ collective action ideas. Pundits and American policymakers often refer to allied "free-riding." Treating alliances as a public good beset by collective action problems generates concern that the United States is "being taken advantage of." US policymakers often use free-riding to criticize lackluster allied defense expenditures. For example, Barack Obama complained in 2016 that "Free riders aggravate me" and US allies "have to pay your fair share." Donald Trump has implied the United States would not protect allies who spend too little on defense. Such complaints and exhortations go back as far as the Eisenhower

administration (Lanoszka, 2015).

Third, increasing great power competition makes assessing Olson and Zeckhauser's argument important for policy debates. Competing visions of US grand strategy hinge in part on the explanatory power of the public goods model. Advocates of retrenchment and "restraint" use the public goods logic to claim that US allies free-ride, so the United States should withdraw from many alliances (Preble, 2009; Posen, 2014). Others assert that alliances do not provide a public good and the benefits of alliance participation outweigh the costs (Brooks, Ikenberry and Wohlforth, 2013; Brands and Feaver, 2017).

The prominence of the public goods model belies a major problem. Fifty-three years after the publication of "An Economic Theory of Alliances," there is little reliable and general evidence about Olson and Zeckhauser's prediction that small states are prone to free-ride. Here, I attempt to fill that empirical gap in the literature.

My research design addresses two key limitations in existing tests of the public goods logic. First, many empirical estimates of free-riding within alliances have a model specification problem with the dependent variable. Olson and Zeckhauser use military spending as a share of GDP to measure contributions to the alliance, and GDP to measure state size. This approach is widely emulated, but it is problematic because GDP is part of the independent and dependent variables. Changes in GDP shift the defense burden, and this deterministic component affects correlation and regression estimates.[1]

One notable paper addresses the dependent variable specification problem, but may not produce general findings. Plümper and Neumayer (2015) examine how growth in military spending by North Atlantic Treaty Organization (NATO) members responds to changing US and Soviet spending. They demonstrate that NATO members are unresponsive to US and Soviet military spending, and present this as evidence of free riding. Plümper and Neumayer (2015) also find no correlation between NATO member size and the extent of free-riding, however, which they argue contradicts Olson and Zeckhauser (1966). This paper's focus on NATO brings me to the second limitation: a

[1] See the appendix for a formal demonstration of this claim.

lack of generalizability.

NATO is the epicenter of free-riding discussions. Following Olson and Zeckhauser's emphasis on military spending as a share of GDP, accusations of free-riding emphasize that NATO members have lower defense burdens than the United States. Scholars, pundits and policymakers have spent decades arguing over how well the public goods model applies to NATO, e.g. (Pryor, 1968; Sandler and Forbes, 1980; Palmer, 1990; Hilton and Vu, 1991; Boyer, 1993; Gates and Terasawa, 1992; Sandler and Hartley, 2001; Lanoszka, 2015; Plümper and Neumayer, 2015; Kim and Sandler, 2019).

Most studies of the public goods model focus on NATO, but NATO is a difficult case for making general conclusions. NATO is exceptionally large, durable and capable. There are only seven tests of the public goods model outside of NATO, most of which examine one or two alliances (Russett, 1970; Starr, 1974; Reisinger, 1983; Thies, 1987; Conybeare and Sandler, 1990; Oneal and Whatley, 1996; Siroky, 2012). Six of these studies estimate correlations between GDP and defense burdens, so they suffer from the aforementioned specification problem. This leaves a need for a general examination of the public goods logic.

Using a Bayesian model that estimates the association between economic weight and percentage changes in military spending within many alliances, I find little evidence that small states are more inclined to free-ride than their larger allies. Given established theoretical skepticism of the public goods model e.g. (Palmer, 1990; Gates and Terasawa, 1992; Sandler and Hartley, 2001; Norrlof, 2010; Niou and Zeigler, 2019), what do these findings add? Some existing theoretical skepticism of the public goods model is motivated by inconsistent correlations between GDP and defense burdens. These correlations do not provide reliable evidence, however. Without a reliable and general test, theoretical revisions of a parsimonious public goods model may be premature.

The chapter proceeds as follows. First, I show why measuring correlations between GDP and defense burdens is likely to produce spurious results. Then I summarize the public goods theory of alliances and use it to derive observable implications of free-riding. Then I describe the model and results. In the final section, I discuss some implications of the findings for scholarship and policy.

2.1 Model Misspecification: GDP and Defense Burdens

Models and correlations between GDP and defense burdens are misspecified. Because defense burdens include GDP in the denominator, changes in GDP affect the expected value of defense burdens. When GDP shifts, the defense burden remains constant only if military spending also changes in such a way that defense spending's share of GDP remains the same. Such changes are highly unlikely.

The deterministic component in the relationship between GDP and defense burdens has important consequences for correlation and linear regression estimates. For a correlation ρ between two variables X and Y:

$$\rho = \frac{Cov(X,Y)}{Var(X)Var(Y)} \tag{2.1}$$

In a linear regression with one independent variable, the coefficient β_1 is equal to:

$$\beta_1 = \frac{Cov(X,y)}{Var(X)} \tag{2.2}$$

Correlations and regression coefficients depend on the covariance between the two variables. In general, covariance is equal to:

$$Cov(X,Y) = E[X,Y] - E[X]E[Y] \tag{2.3}$$

Now, consider the covariance between GDP and military spending as a share of GDP, $\frac{ME}{GDP}$.

$$Cov\left(GDP, \frac{ME}{GDP}\right) = E\left[GDP, \frac{ME}{GDP}\right] - E[GDP]E\left[\frac{ME}{GDP}\right] \tag{2.4}$$

The approximate expected value of the defense burden[2] is equal to:

$$E\left[\frac{ME}{GDP}\right] = \frac{E[GDP]}{E\left[\frac{ME}{GDP}\right]} - \frac{Cov(GDP, ME)}{\left(E\left[\frac{ME}{GDP}\right]\right)^2} - \frac{Var(ME)E[GDP]}{\left(E\left[\frac{ME}{GDP}\right]\right)^3} \tag{2.5}$$

[2]Based on: `http://www.stat.cmu.edu/~hseltman/files/ratio.pdf`.

Substituting this expectation into the covariance of GDP and defense burdens:

$$Cov\left(GDP, \frac{ME}{GDP}\right) = E\left[GDP, \frac{ME}{GDP}\right] - E[GDP]\left(\frac{E[GDP]}{E\left[\frac{ME}{GDP}\right]} - \frac{Cov(GDP, ME)}{\left(E\left[\frac{ME}{GDP}\right]\right)^2} - \frac{Var(ME)E[GDP]}{\left(E\left[\frac{ME}{GDP}\right]\right)^3}\right) \quad (2.6)$$

This then simplifies to:

$$Cov\left(GDP, \frac{ME}{GDP}\right) = E\left[GDP, \frac{ME}{GDP}\right] - \frac{(E[GDP])^2}{E\left[\frac{ME}{GDP}\right]} - \frac{E[GDP]Cov(GDP, ME)}{\left(E\left[\frac{ME}{GDP}\right]\right)^2} - \frac{Var(ME)(E[GDP])^2}{\left(E\left[\frac{ME}{GDP}\right]\right)^3} \quad (2.7)$$

Therefore the expected value of GDP, $E[GDP]$, affects the expected value of defense burdens $E\left[\frac{ME}{GDP}\right]$. This change in the expected value of defense burdens impacts the covariance between GDP and defense burdens, which then informs regression and correlation estimates. Because defense burdens are a non-linear function of military spending and GDP, the consequences of this specification problem for estimates are hard to predict, especially when GDP and military spending are correlated.

The implications of using a ratio variable like defense burdens for inference depend on how GDP and defense spending are correlated. At the very least, the GDP and defense spending values that Olson and Zeckhauser, as well as other researchers e.g.(Oneal, 1990; Kim and Sandler, 2019) employ in cross-sectional correlations depend on past GDP and military spending values. Temporal autocorrelation could generate correlations between GDP and military spending within units. If GDP and military spending decisions are correlated for other reasons, this further complicates the ratio calculations.

In what follows, I report the results of three simulation analyses to assess how cross-sectional correlations between GDP and defense burdens might compare to correlations between GDP and military spending. This simulation is a rough approximation of common tests of burden-sharing within NATO, which estimate annual correlations between GDP and defense burdens among NATO

members. In all three simulations, I simulate an outcome and independent variable, both of which have strong temporal autocorrelation. I then analyze cross-sectional associations between the independent variable and the outcome with OLS. I compare inferences from measuring the outcome as a share of the independent variable to inferences without the ratio. I set the values of the outcome and independent variable to roughly match the values of military spending burdens. The data-generating process contains 100 temporal observations of 20 units. Therefore, I run 100 cross-sectional analyses and each linear regression estimate has 18 degrees of freedom after estimating a regression coefficient and intercept.

In the first simulation, I assume that the independent variable and outcome are entirely uncorrelated, except for spurious associations from temporal autocorrelation. In the second simulation, I assume that the independent variable and outcome are correlated and each variable has temporal autocorrelation. I set the assumed correlation in the second simulation equal to the observed correlation between GDP and military spending in my data, which is .48. Because these analyses follow existing research designs by using cross-sectional snapshots, I do not control for temporal autocorrelation within units. The third simulation uses the simulated data from the second simulation, but converts the outcome into percentage changes, which is the outcome I use in the paper.

Inferences about the association between the simulated independent variable and its ratio with a dependent variable depend on whether the two variables are correlated, as Figure 2.1 shows. This figure plots whether the coefficient in a regression of simulated ratio and non-ratio outcomes on a simulated independent variable is statistically significant at conventional levels. When the independent variable and outcome are independent, creating a ratio generates an overwhelming number of statistically significant findings. When the independent variable and dependent variable are correlated, the ratio variable has a different effect. Creating a ratio of two correlated variables reduces the probability of a statistically significant result, relative to non-ratio estimates. Last, comparing the percentage change measure to a ratio outcome, the ratio outcome is more likely to produce statistically significant regression coefficients. Based on a chi-squared test, there are statistically significant differences in inferences from ratio and non-ratio outcomes for both types

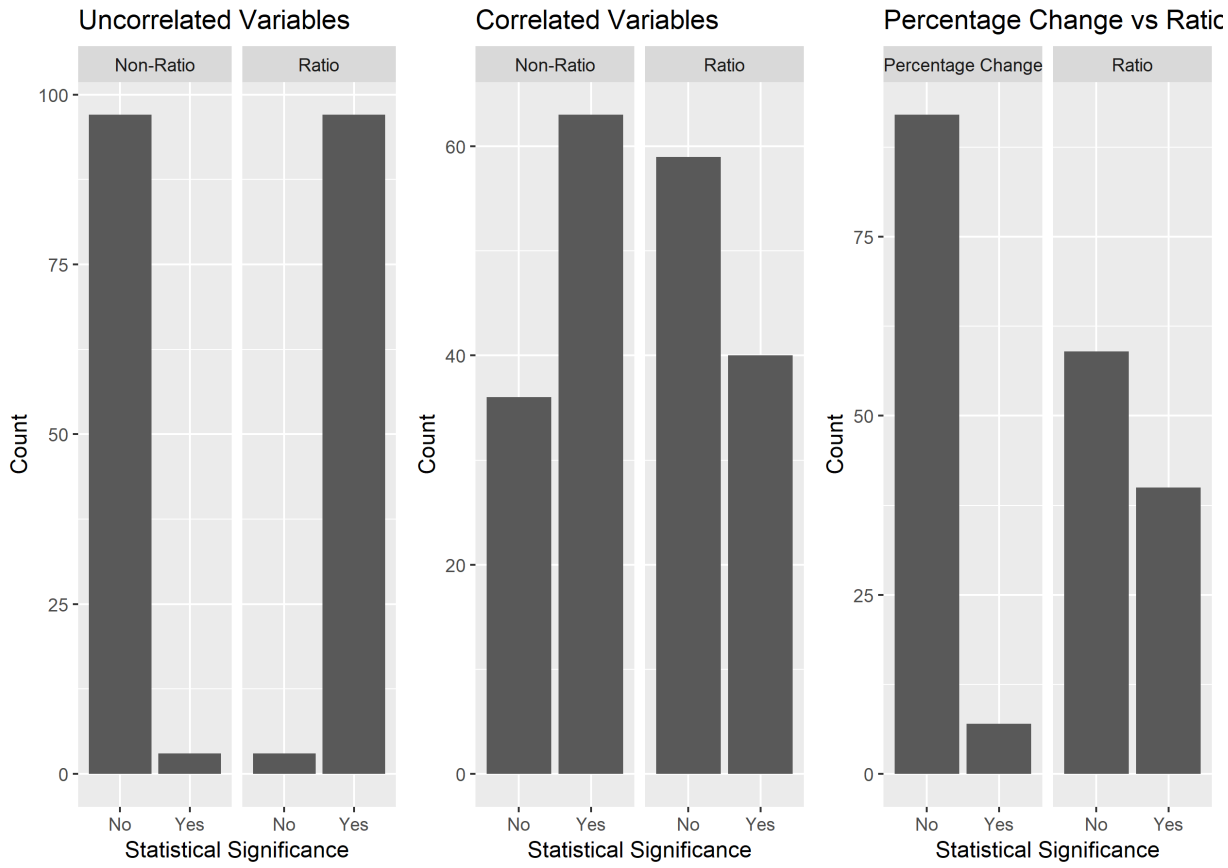

Figure 2.1: Estimated statistical significance of a linear regression coefficient in simulated cross-sectional analyses of ratio and non-ratio variables.

of simulated variables.

In summary, creating a ratio of the dependent and independent variable is likely to produce spurious results. If the two variables are uncorrelated, then estimates may conclude the two variables are correlated, when any association is driven entirely by changes in the independent variable. But if the two variables are correlated, ratio outcomes are more likely to produce findings that do not meet conventional statistical significance thresholds. Last, ratio variables are more likely to produce statistically significant estimates than the percentage changes variable this manuscript and Plümper and Neumayer (2015) prefer. Therefore, estimates of the correlation between GDP and military spending as a share of GDP may produce misleading findings, including null results when GDP and military spending are correlated. Having established this issue, I now generate alternative

predictions of how economic weight might lead to free-riding in alliances.

2.2 Free-Riding in Alliances

To identify observable implications of free-riding, I use the public goods argument to derive predictions about economic size and military spending in alliances. This theoretical exercise does not critique the public goods model. Instead, I take the public goods logic as given, but employ another measure of defense effort. Following Plümper and Neumayer (2015) I use percentage changes in military spending to measure defense effort. Like Olson and Zeckhauser, I conceptualize state size using economic weight.

Why might alliances suffer from a collective action problem? As the aggregate military capability of an alliance provides security for members, states contribute security by investing in their military. Because an alliance cannot exclude members without undermining its purpose, alliance security is a public good. Alliance members receive security regardless of their individual contribution.[3] Thus, states have incentives to rely on their alliance partners and reduce their own military expenditures.

Olson and Zeckhauser expect that larger members of the alliance will invest more in defense, because these states value security from the alliance more. Small alliance members rely on larger partners for security and reduce their defense burdens. As a result, smaller states free-ride on larger alliance participants. Moreover, small states have greater bargaining leverage, because a large state cannot credibly threaten to reduce their contribution and has "relatively less to gain than its small ally from driving a hard bargain" (Olson and Zeckhauser, 1966, pg. 274).

Olson and Zeckhauser compare alliance member size using economic resources, specifically GDP. Economic weight within an alliance, or each state's share of total allied GDP, is a related way to conceptualize differences in state size.[4] Using economic weight facilitates state size comparisons across diverse alliances. A greater share of total economic resources in an alliance gives a state more economic weight and increases their potential military spending contribution.

[3]The marginal costs and benefits of participation depend in part on alliance size, but Olson and Zeckhauser's model shows free-riding even in a bilateral alliance.

[4]A state may be large or small depending on the alliance.

Olson and Zeckhauser expect that economic weight shapes defense expenditures. Smaller members can free-ride and lower military spending. Because larger states place higher absolute value on security from an alliance, alliance participation will increase their investment in military capability, especially if smaller partners reduce defense spending. Therefore, alliance participation will decrease military spending by small states and increase military spending by large states.

> HYPOTHESIS 1: For states with a small share of the total GDP in an alliance, alliance participation will decrease annual percentage changes in military spending.

> HYPOTHESIS 2: For states with a large share of the total GDP in an alliance, alliance participation will increase annual percentage changes in military spending.

Though my predictions use slightly different variables, those variables facilitate a reliable and general empirical test. I test the hypotheses in a sample of all non-microstates from 1919 to 2007.[5] State-year observations are the unit of analysis. To assess Hypotheses 1 and 2, I examine how many alliances have a positive economic weight parameter, and I now describe the test.

2.3 Testing the Public Goods Logic

I examine the two hypotheses by using a Bayesian model to estimate alliance-specific effects of differences in economic weight.[6] This approach provides general evidence about the association between economic weight and military spending across 204 alliances. I use Bayesian estimation because it regularizes estimates with many parameters through partial pooling of alliance parameters.[7]

For each of the 204 ATOP offensive and defensive treaties (Leeds et al., 2002), I estimate the association between economic weight and military expenditures. An economic weight parameter for each alliance captures the consequences of alliance participation. I code economic weight such that a positive economic weight parameter implies more military spending for large members

[5] Limited GDP data makes constructing economic weights for each alliance difficult before 1919. I also omit some alliances after 1919 due to a lack of GDP data.

[6] I also make similar inferences by regressing percentage changes in military spending on a state's average weight in their alliances. See the appendix.

[7] I fit the following model using STAN (Carpenter et al., 2016).

and less spending by small members. A negative economic weight parameter implies that larger members spend less on the military, and small members spend more.

The model starts with state-year percentage changes in military spending y_{it}, transformed with an inverse hyperbolic sine. I model this variable using a t-distribution with degrees of freedom ν to account for heavy tails. The expected value of military spending μ_{it} depends on a constant α, state and year varying intercepts, and control variables $\mathbf{X_{it}}\beta$.[8] σ captures unexplained variation in y.

$$y_{it} \sim student_t(\nu, \mu_{it}, \sigma) \tag{2.8}$$

$$\mu_{it} = \alpha + \alpha^{st} + \alpha^{yr} + \mathbf{X_{it}}\beta + \mathbf{Z_{it}}\gamma \tag{2.9}$$

The $\mathbf{Z_{it}}\gamma$ term captures the impact of economic weight in alliances. $\mathbf{Z}$ is a matrix of state participation in alliances— columns are alliances, rows are state-year observations. If a state is part of an alliance, the corresponding element in $\mathbf{Z}$ depends on their share of total GDP in the alliance. I assigned small alliance members a value of negative one if their economic weight is less than .5 in bilateral alliances, or less than .08 in multilateral alliances. Large alliance members have a value of positive one in $\mathbf{Z}$ if their economic weight is above those thresholds. Having more than half of GDP in a bilateral pact is a clear threshold for asymmetry. I set the threshold to the third quartile of .08 in multilateral alliances because the distribution of economic weight is very different in those alliances. Increasing the multilateral threshold would classify even the US in NATO as a small state, so .08 is a reasonable value. If a state is not part of the alliance, the corresponding matrix element is zero. Multiplying a positive γ by negative one for small states will reduce military spending growth. Multiplying the same positive economic weight parameter by positive one will increase military spending growth for large states.

$\mathbf{Z}$ is a quasi-spatial approach to capturing the impact of participation in multiple alliances.

[8] See the appendix for a full description of all the variables in the model.

In this model, alliance participation affects military spending through economic weight. The γ parameters capture the correlation between economic weight in an alliance and military spending.

γ is a vector of 204 alliance-specific economic weight parameters. Because $\mathbf{Z}$ contains each state's share of allied GDP, these coefficients estimate the association between economic weight and military spending.[9] When a state is not in an alliance, the corresponding γ is multiplied by zero, and has no impact.

Each alliance has a separate impact on military spending, but the economic weight parameters have a common prior distribution where $\gamma \sim N(\theta, \sigma_{all})$. Partial pooling estimates the dispersion of the alliance parameters from the data, so the prior for γ is normally distributed with mean θ and variance σ_{all}. θ is the mean hyperparameter of the alliance coefficients and each γ deviates from θ based on a variance hyperparameter σ_{all}. Every economic weight parameter holds the impact of other treaties constant. A positive γ will lead to changes in military spending growth that match Hypotheses 1 and 2.

Before turning to the results, I briefly summarize the variables in this model. Olson and Zeckhauser use GDP to measure state size, so I constructed a measure of GDP using data from the Maddison Project, which provides longer historical coverage (Bolt et al., 2018). I use military spending data from the Correlates of War Project (Singer, 1988). All alliance membership data comes from Version 4 of the Alliance Treaty Obligations and Provisions (ATOP) data (Leeds et al., 2002).

The dependent variable is percentage changes in military spending. Olson and Zeckhauser use defense spending as a share of GDP as their dependent variable, which is the source of previously described model specification problems (Plümper and Neumayer, 2015). I use percentage changes instead of the defense burden because this measure gives a sense of burdens from changing defense budgets, but has a lower risk of spurious inferences. Annual percentage changes in spending is the change in military spending as a share of the previous year's budget:

[9]This assumes symmetric effects across small and large states, which I relax in the appendix with a weighted coding of $\mathbf{Z}$.

$$\% \text{ Change Military Spending} = \frac{\text{Mil. Ex.}_t - \text{Mil. Ex.}_{t-1}}{\text{Mil. Ex.}_{t-1}} = \frac{\Delta\text{Mil. Ex.}}{\text{Mil. Ex.}_{t-1}} \tag{2.10}$$

Measuring percentage changes in spending matches Olson and Zeckhauser's emphasis on how alliance participants allocate resources to the military. Positive percentage changes in spending imply an expanding defense budget and higher defense burden, all else equal. Moreover, using percentage changes in spending mitigates the risk of spurious inferences due to non-stationarity in panel data. The log-level of military spending is not mean-reverting in long panels. A differenced military spending variable has increasing variance over time, as budgets expand and generate larger changes. Modeling the DV in levels or changes might lead to spurious inferences (Granger and Newbold, 1974).

Using percentage changes in military spending as the dependent variable benchmarks changes to budget size. This facilitates comparisons across states and years. A 2% change is an equally burdensome increase in the defense budget for large and small states, all else equal.

Besides the economic weight values in $\mathbf{Z}$, I controlled for other variables that are correlated with alliance participation and military spending. I adjusted for international war (Reiter, Stam and Horowitz, 2016), civil war participation (Sarkees and Wayman, 2010), and a count of annual MIDs (Gibler, Miller and Little, 2016). I also included measures of regime type, external threat (Leeds and Savun, 2007), GDP, and the Cold War era.

2.3.1 Results

The public goods model predicts many positive economic weight parameters. Because I employed Bayesian modeling, each γ has a posterior distribution.[10] I focus interpretation on the posterior mean and 90% credible intervals.[11] The posterior mean is the expected value of γ, while the credible intervals capture uncertainty around that estimate.

There are no alliances with a clear positive economic weight parameter. Figure 2.2 plots the

[10] See the appendix for a full summary of priors, convergence and model fit. I also show that the model recovers known parameters from simulated data.

[11] I use 90% credible intervals because inferences around 95% intervals are unstable.

γ parameter for each alliance against the start year of the treaty. Points mark the posterior mean. The error bars encapsulate the 90% credible interval.

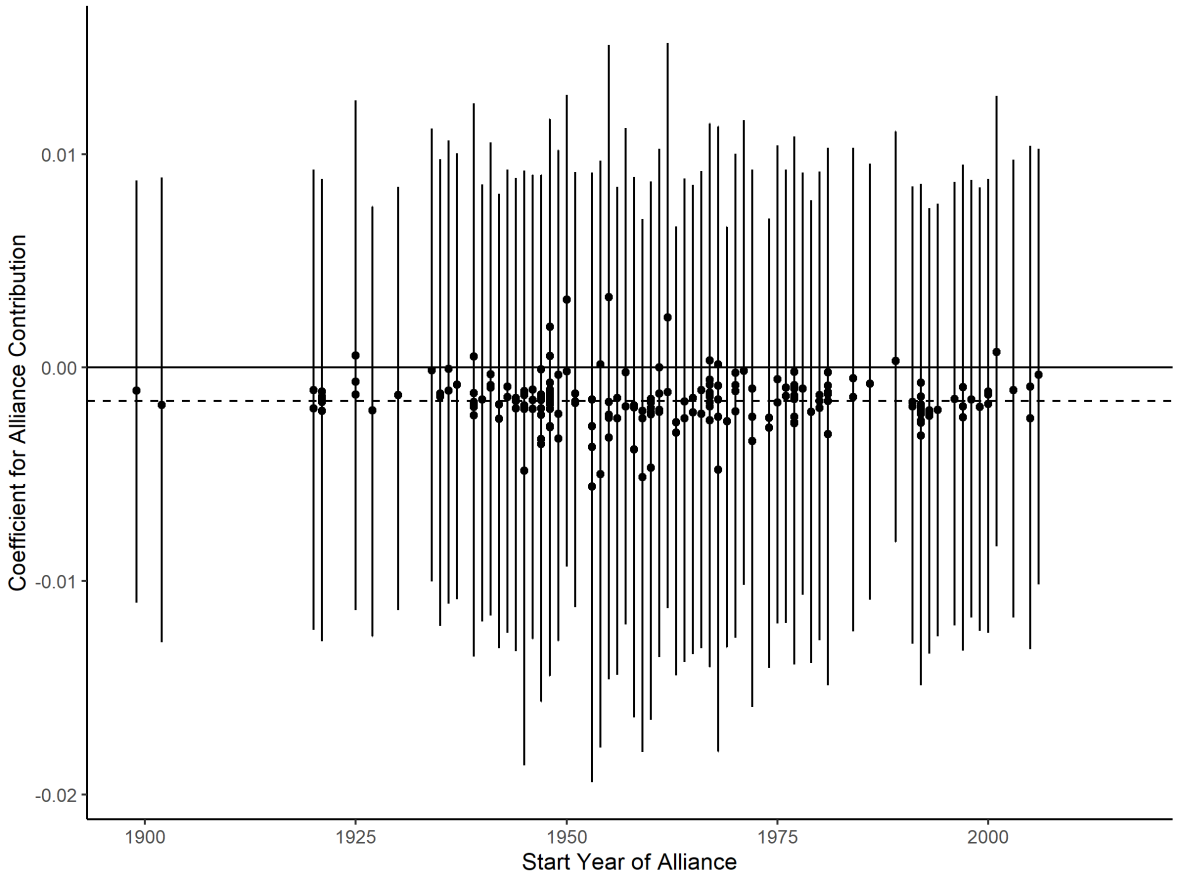

Figure 2.2: Estimated association between economic weight and defense spending growth in 204 defensive and offensive alliances from 1919 to 2007. Positive estimates match the predictions of Hypotheses 1 and 2. Points represent the posterior mean and the error bars cover the 90% credible interval. The dashed line marks the posterior mean of the θ parameter, which is the average association between economic weight and percentage changes in military spending.

No alliances have a uniformly positive 90% credible interval. Most credible intervals are consistent with military spending growth between -0.02 to 0.015. Only 13 of the 204 alliances have a positive posterior mean.

To further examine whether increasing a state's share of allied GDP leads to higher defense spending, I simulated the effect of changing economic weight on percentage changes in military

spending. In the simulated data, I used the full posteriors of the intercept α, all the β coefficients, and one γ parameter. I selected the economic weight parameter with the most positive posterior mass, so this is the *best case alliance* for Hypotheses 1 and 2. I then set the state-level variables at their median or modal value and changed economic weight from -1 to 1.

In Figure 2.3, I summarize predicted changes in military spending at the two economic weight values. In this figure, the point marks the mean and the error bars summarize the 90% credible interval. There is limited evidence that larger alliance participants have higher military spending.

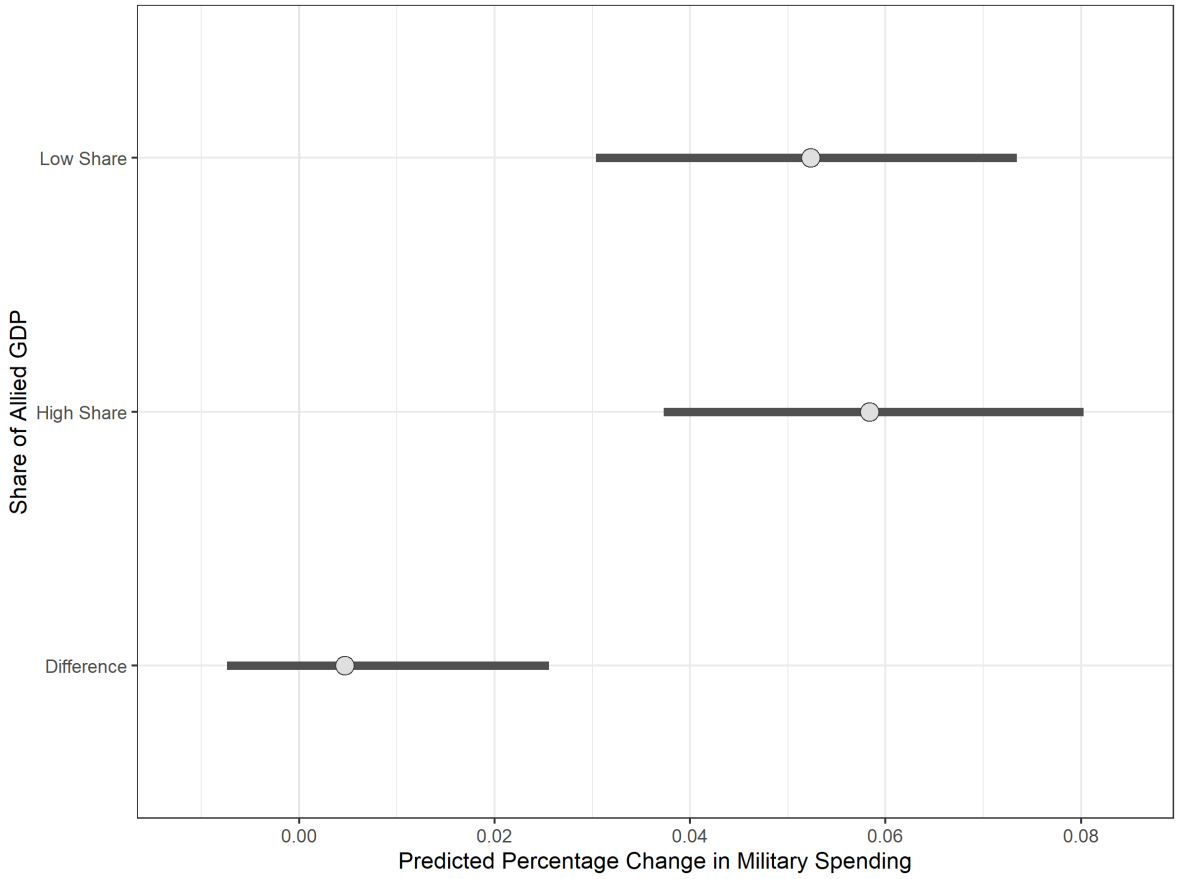

Figure 2.3: Predicted percentage changes in military spending for a simulated state with low or high shares of total allied GDP. Hypotheses 1 and 2 would expect a positive difference between the high and low economic weight scenarios. Points mark the median value, and the error bars summarize the 90% credible interval. The difference estimate captures the effect of moving from low to high economic weight.

The γ estimates also produce inferences about individual alliances. The estimated γ for NATO offers no support for the public goods theory of alliances. The posterior mean of this parameter is -.003, and it has 81% *negative* posterior mass. This finding corroborates the size result of Plümper and Neumayer (2015), but NATO members may still spend less on the military thanks to allied capability (George and Sandler, 2017), rather than economic weight. Among other alliances, the Arab League (ATOPID 3205) has 78% positive posterior mass, which is the most of any alliance. To give another example, the US-South Korea alliance (ATOPID 3240) has 85% negative posterior mass.

2.4 Conclusion

Few alliances see large divergences in military spending by economic weight. Although Olson and Zeckhauser's model is parsimonious, there is little evidence of free-riding based on economic weight. These results provide general evidence that economic weight is not a determinant of free-riding, which matches the finding of Plümper and Neumayer (2015) in NATO. The results do not rule out that alliances provide public goods, or that alliance participation reduces small states' military spending, however.

If alliances do provide public goods, they may only do so in particular circumstances. Small states may only reduce military spending if they believe allied commitments are credible (Goldstein, 1995; DiGiuseppe and Poast, 2016), or leaders are inclined to lower spending (Fuhrmann, 2020). If that is the case, inquiry should emphasize sources of leverage in bargaining between alliance members (Morrow, 1991; Norrlof, 2010; Brooks, Ikenberry and Wohlforth, 2013; Johnson, 2015; Kim, 2016). Alternatively, the extent to which security from an alliance is a public good may vary with factors like technology and strategic doctrine (Sandler and Hartley, 2001).

Beyond scholarship, the results have implications for the inclination of policymakers and pundits use "free-riding" to describe alliance politics. Free-riding is inextricable from a public goods understanding of alliances. But without a clear sense of when or whether alliance provide public goods, free-riding is an inaccurate description of reduced defense effort by alliance participants. Low defense effort could reflect cheap-riding on allied capability, or efficiency gains from specia-

lization in pooled military resources.

Despite these results, collective action remains a central concept in international politics. It may be inappropriate to use alliances to understand collective action problems in international organizations more generally, as Olson and Zeckhauser (1966, pg. 266-7) advocate, but collective action applies to other international organizations. Still, with little evidence of free-riding based on economic weight, policymakers and scholars should be cautious about relying on the public goods model to understand alliance politics.

3. ALLIANCE PARTICIPATION, TREATY DEPTH AND MILITARY SPENDING

Scholars of international relations have long acknowledged that there are two ways for states to increase their security. They can invest in indigenous military capability or form alliances (Morgenthau, 1948; Altfield, 1984; Morrow, 1993). Because both policies provide security, broadly defined, alliance participation should change how states invest in military capability. But exactly how alliances influence military spending remains unclear.

Existing scholarship contains contradictory theoretical predictions and evidence on the question of alliance participation and military spending. One view expects that alliance participation will reduce military spending e.g., (Barnett and Levy, 1991; Morrow, 1993; Conybeare, 1994). The other predicts that alliance participants will spend more on defense e.g., (Diehl, 1994; Morgan and Palmer, 2006; Quiroz Flores, 2011). This paper addresses the divide by using alliance treaty design to explain when alliance participation leads to more or less defense spending. In doing so, it helps clarify a longstanding debate about alliance politics.

I use variation in alliance design and membership to predict how alliance participation affects military spending. Scholars have engaged in extensive study of the sources and consequences of differences in alliance treaty design and membership (Mattes, 2012*b*; Benson, 2012; Poast, 2019*a*; Morrow, 1991; Leeds, 2003*b*; Leeds and Anac, 2005; Fordham, 2010; Mattes, 2012*b*; Poast, 2013; Johnson, Leeds and Wu, 2015). Despite the importance of different alliance treaty designs for outcomes like conflict (Leeds, 2003*b*; Benson, 2012) and trade (Long, 2003; Long and Leeds, 2006) the debate about alliance participation and military spending has largely treated alliances as homogeneous.[1] But alliance participation could plausibly increase or decrease defense expenditures, under different alliance treaties. In particular, I emphasize how treaty depth modifies the impact of alliance participation on military spending. Deep alliances formalize extensive defense cooperation between members. In addition to commitments of military support, deep treaties require defense coordination and cooperation among alliance members.

[1] See DiGiuseppe and Poast (2016) for an important exception.

To explore the consequences of deep and shallow alliances, I examine a particular set of states: non-major powers. I focus on non-major powers because these states clearly show a tradeoff between reassurance and allied military spending in deep alliances.[2] Participation in deep alliances allows non-major powers to reduce military spending due to greater treaty credibility and reduced allied leverage on defense spending. Joining a shallow alliance often increases non-major power military spending because realizing foreign policy gains from alliance participation depends on defense spending when members fear abandonment.

I employ a novel research design to test my argument. First, I develop a latent measure of alliance treaty depth. I then incorporate that measure into a multilevel model which estimates how alliance characteristics modify the impact of total allied defense expenditures on annual percentage changes in military spending. Because allied capability is the ultimate source of security in alliances, I use total allied capability to measure alliance participation. Allied capability is a useful proxy for alliance participation because it combines the impact of joining an alliance and changing allied capability during treaty membership, both of which change the security benefits of alliance participation. Multilevel modeling matches my conditional argument and captures heterogeneous effects of alliance participation across individual treaties. I fit the model on a sample of non-major power states from 1816 to 2007 and find that while deep alliances decrease percentage changes in non-major power military spending, participation in shallow alliances often increases spending.

The argument and findings illuminate a salient debate in US foreign policy about the costs and benefits of alliances. Advocates of deep engagement (Brooks, Ikenberry and Wohlforth, 2013) and restraint (Posen, 2014) in grand strategy have different views of alliances. Proponents of restraint argue that the United States should withdraw from many alliances, because allies spend too little on defense, which then increases US defense spending (Preble, 2009). The deep engagement school argues that the benefits of alliances exceed the costs and believe that the problem of low allied military spending is overstated (Brands and Feaver, 2017). My argument and findings suggest that policymakers can adjust alliance commitments to reassure partners or encourage higher allied

[2] I explore the process connecting alliance participation and military spending for major powers in a separate paper.

military spending, but will struggle to do both. The United States often uses deep alliances to reassure partners, but this may encourage lower allied military spending. This tradeoff suggests that there is little room for compromise between deep engagement and restraint in grand strategy.

The paper proceeds as follows. First, I summarize competing claims on alliance participation and military spending. Then I describe my argument in more detail. After the argument, I present the research design and results. The final section concludes with a discussion of the results and implications for scholarship and policy.

3.1 Do Alliances Increase or Decrease Military Spending?

Scholarship on alliance participation and military spending is divided between two views. One expects that alliance participation will decrease military spending, while the other anticipates a positive relationship between alliance participation and defense expenditures. Each view predicts a different average effect of alliance participation by emphasizing one aspect of alliance politics.

Two types of arguments predict a negative association between alliance participation and defense spending. First, the economic theory of alliances (Olson and Zeckhauser, 1966) claims that alliances are subject to a collective action problem because security from an alliance is a public good. Because alliance security is neither rivalrous nor excludable, members contribute inadequate resources to collective defense. Alliance members can "free-ride" and smaller states exploit larger partners. Lower spending allows alliance members to consume more non-defense goods, but the alliance provides suboptimal security.[3] Second, substitution arguments recognize that states employ one policy in place of another (Most and Starr, 1989). Alliances provide security without requiring additional military spending (Morrow, 1993; Conybeare, 1994). Given extra security, states rely on their allies and and reallocate military spending to other goods. Both the substitution and public goods models expect that alliance participation reduces military spending due to the opportunity costs of military expenditures. States want to rely on their allies for security because higher defense expenditures leave fewer resources for other goods (Fordham, 1998; Fearon, 2018).

[3]Sandler and Forbes (1980), Oneal (1990) and Sandler and Hartley (2001) all modify the public goods logic while relying on Olson and Zeckhauser's core intuition.

A contradictory perspective asserts that alliance participation increases military expenditures. Several arguments predict higher military spending by alliance members, using a shared intuition that states increase military spending to support their alliance commitments. Diehl (1994) argues that alliances create new foreign policy obligations, necessitating extra military spending. Because alliances expand what a state can achieve in international relations, states might increase military spending to pursue other foreign policy goals (Morgan and Palmer, 2006). For example, buffer states use conscription to make themselves a more attractive alliance partner (Horowitz, Poast and Stam, 2017). Others assert that cooperation within alliances generates higher defense spending (Palmer, 1990; Quiroz Flores, 2011). These predictions of a positive correlation between alliance participation and military spending contradict expectations of lower military spending by alliance members.[4]

3.1.1 Mixed Evidence

Mixed findings reinforce theoretical divisions between the contradictory views of alliances. Some studies find a positive association between alliance participation and military spending. Others find a negative relationship.[5]

Scholars have studied the connection between alliance participation and military spending in two ways. General studies of military spending and alliances compare many states through dummy indicators of alliance participation, which collapse alliances into a state-level measure. This design compares states with at least one alliance to those with none. By contrast, specific research designs examine defense expenditures within a single alliance. Much of this work seeks to explain variation in spending among NATO members, focusing on how states respond to increases in contributions to collective defense by the United States.

Table 3.1 summarizes previous results on the issue of alliance participation and military spending. Both specific and general research designs produce mixed findings. There is one negative,

[4] Senese and Vasquez (2008) argue that military spending and alliances are part of a conflict spiral of simultaneous growth in military expenditures and alliance participation, which suggests that conflict behavior drives any correlation between alliances and military spending.

[5] Because tests of the public goods model use military spending as a share of GDP as the their outcome of interest, I do not include most of those results in this summary.

three positive and two null estimates of the correlation between alliance participation and spending in general studies. Specific studies turn up five negative and two positive results.

Research Design	Study	Decrease	Increase	Null
General	Most and Siverson (1987)			X
	Conybeare (1994)	X		
	Diehl (1994)		X	
	Goldsmith (2003)			X
	Morgan and Palmer (2006)		X	
	Quiroz Flores (2011)		X	
Specific	Conybeare and Sandler (1990)		X	
	Barnett and Levy (1991)	X		
	Morrow (1993)	X		
	Sorokin (1994)	X		
	Chen, Feng and Masroori (1996)		X	
	Plümper and Neumayer (2015)	X		
	George and Sandler (2017)	X		

Table 3.1: Findings of the association between alliance participation and military spending. The top block details results from general studies, which compare states with at least one alliance to states without any alliances. The bottom block shows results from specific studies, which examine how national military spending changes in response to shifting allied capability.

The mixed empirical results reflect a theoretical problem. Both perspectives make unconditional claims about the average effect of alliance participation on military spending. With one exception (DiGiuseppe and Poast, 2016), scholarship on alliance participation and military spending ignores differences between alliances. Treaty obligations and membership vary widely across alliances, however (Leeds et al., 2002). I focus on a key difference between alliances that can help us understand their heterogeneous effects on military spending: the depth of military cooperation in the treaty.

3.2 Argument

My argument explores how deep military cooperation in an alliance treaty modifies the impact of alliance participation on non-major power military spending. Because non-major powers

focus on security, alliances protect them from external threats. Treaty depth shapes whether the security benefits of alliance participation depend on military spending. Even in the face of significant threats, non-major powers can gain more security with less military spending if they form deep alliances. Conversely, security gains in shallow alliances often depend on non-major powers maintaining or even increasing their military spending to manage abandonment concerns.

I focus on non-major powers to maintain theoretical and empirical parsimony. As I explain below, major and non-major powers use alliances to achieve different foreign policy goals, which changes how alliance participation affects their military spending. My argument also provides novel insights about non-major powers. Some scholarship and much popular discourse assumes that non-major powers regularly reduce military spending in alliances. I challenge this assertion by showing that alliance participation sometimes requires non-major powers to increase defense expenditures.

I start the argument by describing a general framework where alliances are institutionalized military cooperation between states. Then I discuss how deep formal military cooperation affects alliance credibility. Last, I explain how alliance depth affects the connection between alliance participation and non-major power military spending.

3.2.1 Cooperation in Alliances

Alliances are a form of international cooperation. Promising military support through a treaty generates a credible commitment of intervention (Fearon, 1997; Morrow, 2000). Allied support then helps members achieve crucial foreign policy goals like deterrence or winning wars (Walt, 1990; Snyder, 1997). States form alliances so they can use other states' military capabilities to back their foreign policy aims.

Because allied capability gives a treaty foreign policy value, alliance participation is inseparable from allied capability (Fordham and Poast, 2014). The presence of an alliance treaty formalizes when a state can expect military intervention (Morrow, 2000), but the treaty itself does not provide security. Allied capability is the ultimate source of security in an alliance. Greater allied capability increases the value of an alliance, all else equal (Johnson, Leeds and Wu, 2015), so I conceptualize

alliance participation in terms of allied capability.[6]

Alliance treaties and the capability they draw on can support many foreign policy aims, which often facilitates exchanges between alliance participants. One common exchange occurs in asymmetric alliances between major and non-major powers (Morrow, 1991). Large states form asymmetric alliances to increase their foreign policy influence, while smaller partners gain protection from external threats. Not all alliances are asymmetric,[7] but the divergent motives of major and non-major powers in these treaties reflect general tendencies in alliance politics. Major powers often use alliances to address the global balance of power and increase their influence. Smaller non-major powers tend to emphasize immediate security and regional concerns in their alliances. As a result, there are distinct processes behind non-major and major power alliance participation and these states respond to allied capability and treaty design in different ways.

Non-major powers use allied support to provide protection from external threats. But as with all cooperation, alliance members must account for opportunism, or "behavior with guile" (Williamson, 1985). Even as states commit to an alliance, they can also benefit from defecting and taking advantage of allied cooperation. Sometimes the perceived benefits of defection outweigh the long-run benefits of cooperation, so alliance members face an enforcement problem (Fearon, 1998*a*; Koremenos, Lipson and Snidal, 2001).

Non-major powers are especially concerned with abandonment, which is the most common form of opportunism in alliances. One estimate suggests that the rate of compliance with military intervention obligations is only 50% (Berkemeier and Fuhrmann, 2018). For security-focused non-major powers, these alliance violations threaten their main goal. Concern with abandonment then affects military spending decisions.

Abandonment and military spending are related because greater alliance credibility allows states to lower military expenditures.[8] Though states can augment the collective military capability

[6] A binary conceptualization of alliance participation assumes all alliances are equally valuable.

[7] 130 of 289 ATOP alliances with offensive or defensive obligations are asymmetric pacts with at least one major and one non-major power, but a further 122 alliances are symmetric treaties between non-major powers.

[8] The public goods model of alliances calls reduced military spending in alliances free-riding. I do not use this language because it has come to imply that reduced defense spending is problematic. But as Olson and Zeckhauser (1966, pg. 278) themselves observe, lower defense spending is not normatively problematic.

of an alliance through investing in military capability, they can also reduce defense spending and rely on their partners (Olson and Zeckhauser, 1966; Morrow, 1993; Conybeare, 1994; Sandler and Hartley, 2001). Under credible alliances, such reductions in military expenditures are more likely, because states have less fear of abandonment.

As DiGiuseppe and Poast (2016) observe, some alliances have fewer credibility concerns due to members' political regime type. They show that defense pacts with democracies lower defense spending, as democracies make more credible commitments. This insight about conditional credibility is a useful starting point because credibility is multifaceted. In treaty design, depth, unconditional military support (Benson, 2012; Chiba, Johnson and Leeds, 2015) and issue linkages (Long and Leeds, 2006; Poast, 2012, 2013) all increase credibility.[9] In this paper, I focus on treaty depth, which is a costly promise that allows alliance members and potential adversaries to infer the credibility of the alliance (Leeds, 2003*b*; Fuhrmann and Sechser, 2014).

If alliance participation can reduce non-major power military spending, is that a problem for their allies? This argument assumes that allied states usually prefer higher non-major power military spending because it adds capability to the alliance.[10] Allies also encourage higher military spending so their partners can manage internal threats. This preference for higher defense spending is not a driving force in alliance treaty design, however, because states prioritize abandonment and entrapment concerns in treaty design. States cannot form an alliance without addressing abandonment and entrapment, but they can leave bargaining over military spending to alliance maintenance (Snyder, 1997). Also, some evidence suggests that alliance negotiations emphasize when and how states provide military support (Poast, 2019*a*).

Assuming that allies of non-major powers usually prefer higher military spending is reasonable. States often sell or transfer weapons to allies in order to augment allied capability (Yarhi-Milo, Lanoszka and Cooper, 2016; Pamp, Dendorfer and Thurner, 2018). Even in alliances between superpowers and non-major powers, where highly asymmetric capability and interests might reduce

[9]Though the argument emphasizes depth, the research design accounts for multiple sources of alliance credibility.

[10]This preference is somewhat attenuated by concern that substantial changes in capability will provide too much autonomy. For example, the United States has inhibited nuclear acquisition by some allies (Gavin, 2015; Lanoszka, 2015).

concerns about non-major power defense expenditures, superpowers often encourage higher military spending. US demands for higher military spending by NATO members are well-known (Schuessler and Shifrinson, 2019). The Soviet Union also regularly pressured Warsaw Pact members to increase defense spending (Simon, 1985).

When allies want non-major powers to spend more, enforcing such cooperation is difficult. Normative appeals to common interests are ineffective. Though verbal communication or "cheap talk" has value in international politics (Trager, 2010), it is unlikely to overcome the opportunity costs of defense spending. Even after reducing defense expenditures, alliance members retain foreign policy benefits and can reallocate resources to other priorities. The ability to reduce defense spending and spend more on other goods sometimes motivates states to form alliances (Kimball, 2010; Allen and DiGiuseppe, 2013).

Therefore, if alliance members want to encourage greater allied defense spending, they need leverage. States gain leverage either from a credible threat to abandon low-spending partners or direct control over allied policies. Policy control of allied spending decisions occurs when the alliance reflects hierarchical relationships (Lake, 1996). Without such direct influence, states must possess a credible threat to leave the alliance over low defense spending. Otherwise, allies will dismiss weaker signals and threats due to uncertainty and incomplete information.

Reassuring allies reduces the credibility of threats to abandon states that spend too little on defense. States cannot simultaneously reassure their allies and maximize leverage on defense spending. As alliance members use costly commitments to reassure, partners have more freedom to reduce defense spending.

Under less credible alliances, low military expenditures could increase the risk of abandonment, so members have less freedom to reduce defense spending. Moreover, although states in less credible alliances increase their foreign obligations, they also face the risk of having to meet foreign contingencies without allied support. As treaty credibility falls, the foreign policy benefits of alliance participation become contingent on military spending.[11] This occurs because alliance

[11]For example, (Niou and Zeigler, 2019) use a formal model to show that in alliances between rivals, states are less likely to reduce military spending, because their foreign policy gains from alliance participation depend on military

members must hedge against abandonment and partners can leverage that concern to encourage higher defense spending. Thus, alliances with less credibility will be more likely to increase military spending.

Treaty depth highlights this tradeoff between reassurance and military spending. Where deep alliances often reduce non-major power military spending, participation in shallow alliances often increases it. Stipulating deep cooperation reassures partners and reduces leverage over allied military expenditures. Credibility from treaty depth also promotes efficiency gains from coordinated defense effort. I now describe the role of treaty depth in more detail.

3.2.2 Alliance Treaty Depth

I focus on depth because it provides theoretical leverage to predict when alliance participation increases and decreases military spending, which reveals a trade off between reassurance and defense spending. Moreover, treaty depth is a common policy choice.[12] Though states probably do not change their political regime to reassure allies, they often form deep alliance treaties. Over half of defensive or offensive ATOP alliances have some depth.

Alliance depth is the extent of defense cooperation formalized in the treaty. Deep alliances require additional policy coordination and military cooperation beyond a promise of military support. While shallow alliances stipulate more arms-length cooperation between members, deep treaties lead to closer cooperation. Defense cooperation in a deep alliance takes many forms. Allies can form an integrated military command, provide military aid, commit to a common defense policy, provide basing rights, set up an international organization or undertake companion military agreements. All of these obligations move alliance members away from an arms-length partnership towards close cooperation via policy coordination and regular interaction, while imposing monetary and policy autonomy costs.[13]

spending.

[12] In a related paper, I explore the sources of alliance treaty depth. All the sources of depth that paper identifies are included as control variables in the empirical analysis. Average democracy and threat at the time of alliance formation are the two largest sources of higher treaty depth.

[13] Although depth can have monetary costs, reductions in military spending due to greater treaty credibility will often outweigh those costs. Alliance contributions are a small part of most defense budgets, and non-major powers rarely use basing rights in deep alliances to deploy their troops abroad.

One example of a deep alliance is a 1948 defense pact between the United Kingdom and Jordan. This alliance included basing rights, military aid, official military contact, and an Anglo-Transjordan Joint Defense Board. The basing, aid, and cooperation added more depth than a 1912 treaty between Greece and Bulgaria which only commits to mutual defense and consultation if either state is attacked by Turkey.

Deep alliances like the Anglo-Jordanian pact reduce non-major power military spending in two ways. First, depth reassures partners and reduces leverage on defense spending. Deep alliances are more credible because defense cooperation is costly. Making costly commitments of bases, policy coordination, or aid reassures allies (Morrow, 1994). Depth is especially useful because alliance members face a time inconsistency problem. Alliance treaty fulfillment depends on shared foreign policy interests (Morrow, 2000; Leeds, 2003*a*), so changing foreign policy interests threaten alliance fulfillment (Leeds and Savun, 2007). A deep alliance makes a series of repeated transfers, and states can signal commitment by maintaining those transfers.[14] Second, credibility in deep alliances can facilitate more efficient defense spending, as alliance members to provide specific capabilities. Specialization means members of deep alliances spend less on the military but retain adequate security. These efficiency gains depend on the credibility of the alliance, because states will only specialize if they believe the alliance is reliable (Leeds, 2003*a*).

On the other hand, shallow alliances are less credible, so participation in these treaties is more likely to increase military spending. Shallow alliances have some basic credibility from hands-tying signals (Fearon, 1997), as well as the audience (Morrow, 2000) and reputational (Gibler, 2008; Crescenzi et al., 2012) costs of violation. Even so, threats to abandon low-spending allies are more credible than in a deep alliance where partners have taken pains to reassure their partners. In a shallow alliance, members must hedge against abandonment, which partners can use as leverage to discourage low defense spending. Maintaining the benefits of alliance participation then requires defense spending, because low military spending might endanger the treaty or expose states to adverse consequences if they are abandoned. Shallow alliances are also less likely to fa-

[14]Conversely, eliminating or reducing planned transfers reduces the credibility of the whole alliance.

cilitate military specialization due to the fear of abandonment. As a result, participation in shallow alliances often increases military spending.[15]

To illustrate the logic, consider two related alliances from the inter-war period. A 1920 treaty between France and Belgium (ATOPID 2055) added commitments of military aid and policy coordination to defensive obligations. Given this depth, the Franco-Belgian alliance reduced Belgian defense expenditures, even while Belgium helped occupy the Ruhr. A more limited treaty with only military support between France, Belgium, the United Kingdom, Italy and Germany (ATOPID 2130) increased Belgian spending, on the other hand.

Taken together, the examples and argument suggest that treaty depth modifies the impact of alliance participation on non-major power military spending. Shallow alliances often increase military spending, and deep alliances usually reduce spending. Therefore, with a continuous conceptualization of depth, there should be a negative correlation between treaty depth and the impact of alliance participation on non-major power military spending as the positive effects of shallow treaties turn into negative effects in deep alliances. This implies three separate hypotheses, one about shallow alliances, another about deep alliances, and the third about how changes in treaty depth modify the association between alliance participation and military spending.[16]

HYPOTHESIS 1: ON AVERAGE, PARTICIPATION IN SHALLOW ALLIANCES WILL INCREASE PERCENTAGE CHANGES IN NON-MAJOR POWER MILITARY SPENDING.

HYPOTHESIS 2: ON AVERAGE, PARTICIPATION IN DEEP ALLIANCES WILL DECREASE PERCENTAGE CHANGES IN NON-MAJOR POWER MILITARY SPENDING.

HYPOTHESIS 3: AS ALLIANCE TREATY DEPTH INCREASES, THE IMPACT OF ALLIANCE PARTICIPATION ON PERCENTAGE CHANGES IN NON-MAJOR POWER MILITARY SPENDING WILL DECREASE.

[15] One objection to this argument is that deep alliances are more valuable to members, which augments allied influence on defense spending. Although alliance value adds some leverage, it cannot offset reducing the credibility of threats to abandon low-spending allies. Value increases leverage because states fear their allies will abrogate a valuable alliance, and deep alliances counteract this essential concern.

[16] Hypothesis 3 follows from Hypotheses 1 and 2.

Before detailing how I test these hypotheses, I must briefly justify two conceptual and measurement choices. First, the three hypotheses predict how percentage changes in non-major power military spending differ under deep and shallow alliances. Percentage changes in military spending express changes in spending as a share of the previous year's defense budget. This variable is an appropriate outcome of interest, in part because it expresses the opportunity costs of military spending. All else equal,[17] a larger increase in spending relative to the previous year's defense budget imposes more constraints on other goods. Using percentage changes also facilitates comparisons across diverse states and years.

Second, I use allied capability to express the consequences of alliance participation for military spending, rather than a dichotomous indicator of participation. Measuring participation using allied capability matches my argument and encompasses the division in previous research designs. My argument starts with the premise that states form alliances so that allied capability supports their foreign policy goals. States do not respond to a treaty per se, rather they respond to expectations that allies will employ military capability on their behalf. This makes alliances with more capable states are more valuable, all else equal. If allied capability is the main source of security, alliance participation affects military spending through joining an alliance and changes in allied capability after the treaty forms. In previous scholarship, general research designs address the implications of joining an alliance, while specific designs focus on the ramifications of changes in allied capability. Conceptualizing alliance participation in terms of allied capability encapsulates both designs, creating a unified approach to understanding how alliances affect military spending.

Because my argument focuses on differences between deep and shallow treaties, the research design must measure alliance treaty depth and show how depth modifies the impact of allied capability on military spending. I use a measurement model to infer depth from treaty content, then connect alliance characteristics to military spending with a multilevel model. The next section describes the research design in more detail.

[17] Especially holding economic growth constant.

3.3 Research Design

The research design involves two steps. First, I develop a latent measure of treaty depth for alliances with military support. Second, I employ that measure in a multilevel model to estimate how treaty depth modifies the impact of alliance participation on military spending. I then estimate the multilevel model in a sample of non-major powers from 1816 to 2007. The next section describes the measure of alliance treaty depth.

3.3.1 Measuring Alliance Treaty Depth

I start by assuming that formal treaty commitments reflect alliance depth.[18] I then use observed alliance treaty characteristics to infer depth. This approach could produce ordinal or continuous measures of treaty depth. Before describing my measure, I discuss two related measures in previous work, one of which is ordinal, and the other of which is continuous.

In an ordinal index of treaty depth, researchers theoretically assign a range of depth values. (Leeds and Anac, 2005) develop an ordinal measure by assigning alliances military institutionalization scores of zero, one or two based on the extent of investment in joint action required by the alliance treaty. The resulting measure roughly captures treaty depth, but it understates variation in treaty depth. This measurement strategy imposes equal weights on different depth sources and does not aggregate multiple sources of depth. For example, it treats an integrated military command and military bases as equivalent sources of depth, and does not aggregate the two promises if both are present. Given these limitations, I employ a more flexible measurement strategy.

I use latent variable modeling to create a continuous measure of treaty depth that makes more nuanced distinctions between alliances. The measurement model uses correlations between observable alliance treaty content and unobserved latent depth to predict the depth of each treaty. With this approach, theory identifies the relevant correlates of treaty depth, but the data drives how much depth each observed variable adds to the alliance.

Measurement models have a rich history in political science (Clinton, Jackman and Rivers,

[18] Formal treaty obligations may not be fully implemented, but formal depth is more likely to produce practical depth as states try to uphold the credibility of their commitments.

2004; Treier and Jackman, 2008; Fariss, 2014). My particular measure builds on work by Benson and Clinton (2016), who use a latent variable model (Quinn, 2004) to measure alliance scope, depth and capability. I emulate Benson and Clinton's approach, but use a different concept, sample of alliances and estimator. Conceptually, Benson and Clinton (2016) define depth as the general costliness of the alliance, so they include measures of economic issue linkages and secrecy. My definition of depth emphasizes military cooperation only, and I treat issue linkages as a separate source of credibility. Given their broad definition of depth, Benson and Clinton also include neutrality pacts in their sample of alliances. I am only interested in offensive and defensive alliances, however. As for the estimator, in some fully parametric latent variable models, the latent variables influence the form of the dependence structure and the marginal distributions of the latent value estimates. These correlations can produce misleading inferences about latent scores, so I employ a semiparametric estimator.

Due to the limits of ordinal measures and conceptual differences with Benson and Clinton's latent measure, existing measures of treaty depth do not fit my purposes in this paper.[19] Therefore, I create a new measure of treaty depth in offensive and defensive ATOP alliances using a semiparametric factor analysis. I use a Bayesian Gaussian Copula Factor Model (Murray et al., 2013) to measure alliance treaty depth. Murray et al's model improves inferences from mixed factor analysis for continuous, ordinal, and binary observed data by relaxing distributional assumptions. Given discrete observed variables and non-Gaussian latent variables, the dependence among the latent variables and their marginal distributions are both influenced by the latent variables. This approach breaks the dependence between the latent factors and marginal distributions by using copulas to encode the dependence among the latent variables.[20] Beyond the semiparametric aspect, this measurement model is a standard ordinal factor analysis.

I estimated the measurement model using observed data from 289 alliances with offensive or defensive obligations in the alliance-level ATOP data (Leeds et al., 2002). I examine alliances with

[19]See the appendix for a more detailed justification of this choice and evidence that Leeds and Anac (2005)'s measure of institutionalization produces similar inferences.

[20]Copulas are a distribution function on $[0, 1]^p$ where each univariate marginal distribution is uniform on [0, 1].

military support because prior studies of alliance participation and military spending focus on these treaties. Indicators of treaty depth include military aid, bases, international organization formation, integrated military command, defense policy coordination, subordination of forces in wartime, specific contribution requirements, and commitments to form companion military agreements.[21] The argument suggests there is a single factor underlying variation in all eight indicators, so I fit the model with one latent factor. To estimate the model, I used Parameter expanded Gibbs sampling, the default generalized double Pareto (GDP) prior, 20,000 burn-in iterations of the MCMC chain, and 30,000 samples thinned every 30 observations to ensure convergence. The estimates include posterior distributions for the factor loadings and the latent factor.

To summarize treaty depth, I use the posterior mean of the latent factor for each alliance, so each alliance has its own depth value. The posterior mean captures the expected depth of an alliance treaty, conditional on its formal promises. Figure 3.1 describes the latent depth of ATOP alliances with defensive or offensive commitments from 1815 to 2016. There is substantial variation in alliance treaty depth, which has several sources. The top panel in Figure 3.1 shows the factor loadings from the latent variable model, which are essentially correlations between the observed variables and latent factor. Policy coordination, integrated military command, and formal organizations are the three largest correlates of depth. The other five factors have similar associations with latent treaty depth.

The bottom panel of Figure 3.1 plots the posterior means and uncertainty of the depth estimates against the start year of the treaty. Many treaties have no depth, and are clustered on around -0.8. 171 alliances have a depth score higher than -0.6 because at least one source of depth is present. Even after accounting for uncertainty, it is possible to distinguish between some alliances.

Although the values of the latent measure are not intrinsically meaningful, differences between treaties on the latent scale are informative. The median of treaty depth is -0.09, and the mean is 0.05. The median treaty is the Southeast Asian Treaty Organization (SEATO), which includes a formal international organization (ATOP ID 3260). There are many shallow treaties that only

[21] These are the variables Leeds and Anac (2005) use, with the addition of a companion military agreements dummy.

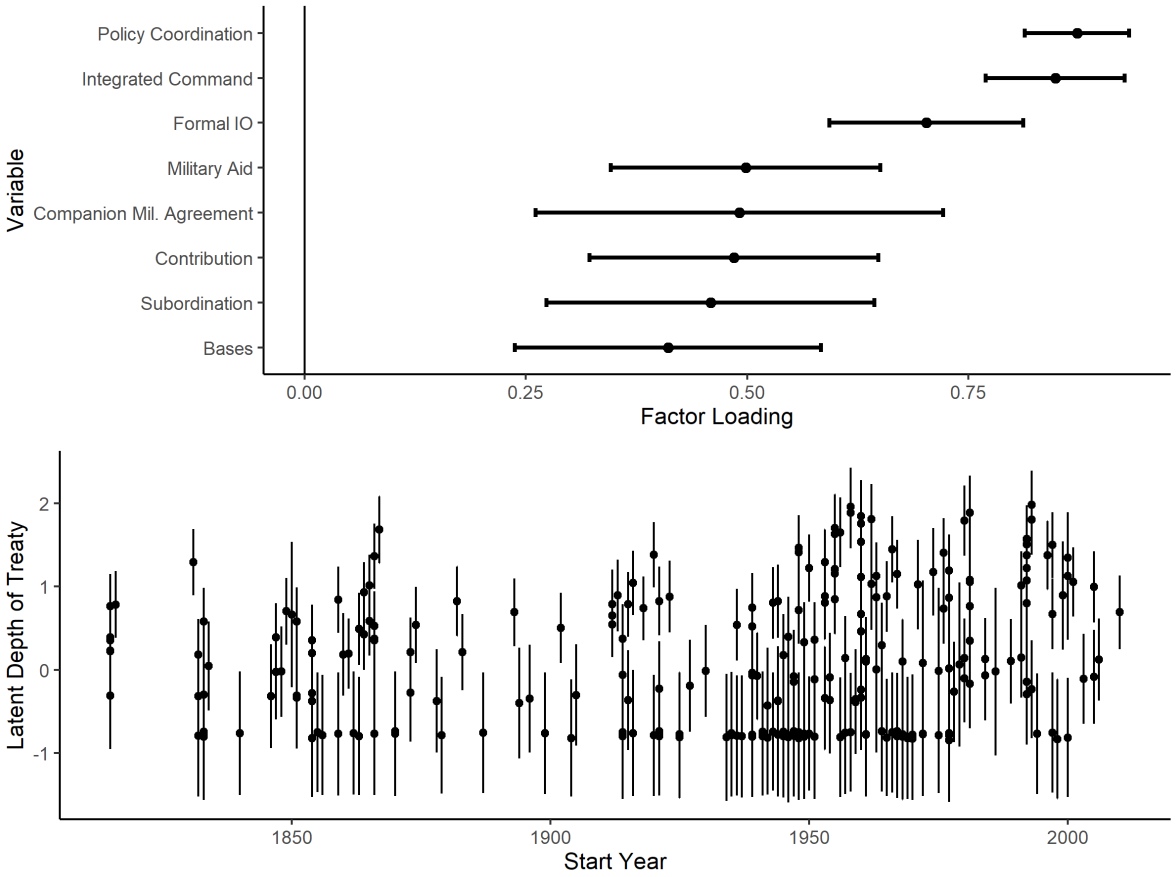

Figure 3.1: Summary of the latent measure of alliance treaty depth for 289 defensive or offensive alliances from 1816 to 2016. The top panel plots the factor loadings with 90% credible intervals. The bottom panel plots mean treaty depth (points) and the standard deviation (error bars) against the start year of the treaty.

include military support. One such alliance is an 1855 pact between France, the UK and Sweden (ATOPID 1190) which promises defense and consultation.

Three of the deepest treaties are a 1993 alliance between Russia and Tajikistan (ATOPID 4470), a 1958 alliance between the UAE and Yemen (ATOPID 3345), and a 1981 pact between Gambia and Senegal (ATOPID 3930). All these alliances stipulate extensive defense cooperation. The alliance between Russia and Tajikistan includes military aid, bases, a companion military agreement, and integrated military command. The other two treaties attempted to establish a federation through military support, international organizations, basing, and defense policy coordination.

The latent measure has some face, concept, and discriminant validity. As an example of face validity, the Gambia-Senegal federation requires deeper cooperation than arms-length commitments of military support. Shallow treaties promise little beyond military support, matching my conceptualization of treaty depth. Last, Figure 3.1 shows that this measure can distinguish between deep and shallow commitments.

This latent treaty depth measure is the key explanatory variable in my empirical analysis. My argument claims that differences in depth among alliances modify the impact of alliance participation on percentage changes in military spending by alliance members at the state-year level of analysis. I use a multilevel model to capture this relationship, and summarize the estimation strategy in the next section.

3.3.2 Estimation: Multilevel Model

Because alliances and state-year observations are separate levels of analysis, I use a multilevel model to estimate the association between treaty depth and military spending. Multilevel modeling bridges levels of analysis (Steenbergen and Jones, 2002; Gelman and Hill, 2007). My model estimates heterogeneous effects of alliance participation on military spending as a function of alliance characteristics, in order to make inferences about how alliance characteristics like formal depth modify the impact of individual alliances on military spending. To facilitate computation and interpretation, I fit the model using Bayesian estimation in STAN (Carpenter et al., 2016).

See the appendix for details of the weakly informative prior distributions and evidence the chains converged.

This research design is more complicated than a panel data model like the estimator DiGiuseppe and Poast (2016) use,[22] but the multilevel components add substantial value, especially by connecting the argument and research design. I argue that treaty depth modifies the impact of alliance participation on growth in military spending. Differences in how alliance participation impacts military spending are the outcome of interest. The multilevel model explicitly compares the impact of participation in deep and shallow alliances by estimating how changes in treaty depth modify the consequences of alliance participation.

Unlike the multilevel model, standard panel models employ state-level proxies for alliance characteristics, which compare states rather than alliances. This practice of aggregating alliances at the state-year level of analysis may produce misleading inferences (McElreath, 2016, pg. 356). Summarizing alliance characteristics at a different level of analysis changes the mean and variance of key independent variables, which could affect inferences. Multilevel modeling avoids this aggregation problem by retaining the structure of alliance data, where states participate in multiple alliances and alliances could have heterogeneous effects on military spending. The multilevel model estimates the specific impact of each alliance on members' military expenditures, thereby revealing differences between individual treaties. On the other hand, aggregating multiple alliances into state level indicators will mask any heterogeneous effects of individual treaties.[23]

Furthermore, multiple alliance characteristics affect the consequences of alliance participation. The multilevel model captures multiple sources of heterogeneity in how alliances impact military spending. Treaty depth is correlated with other aspects of alliance membership and design, so this step is important.[24] In a panel estimator with state-level proxies for alliance characteristics, accounting for correlated alliance characteristics is difficult. To acocout for several alliance characteristics, panel estimates must average different parts of a state's alliance portfolio or subset the

[22] See the appendix for results from several models with state-level indicators of alliance depth.

[23] Partial pooling of the alliance-specific parameters generates reasonable estimates for each alliance.

[24] For example, I show in another paper that democratic alliance membership is positively correlated with treaty depth.

data. Averaging reduces theoretically interesting alliance-level variation, and analysis of multiple subsets risks generating spurious findings through multiple comparisons. In a multilevel model, I can account for how multiple alliance characteristics change the consequences of alliance participation by including other variables besides treaty depth in an alliance level regression. Therefore, my estimate of how treaty depth modifies the impact of alliance participation on military spending holds key alliance and state characteristics constant. I now describe the model specification in more detail.

3.3.2.1 Model Specification

This multilevel model connects two distinct regressions. The base is a state-year-level regression, which includes the impact of alliance participation. A second alliance-level regression modifies the effect of alliance participation on military spending, like an interaction.

The state-year-level regression starts with a distribution for the outcome:

$$y \sim student_t(\nu, \mu, \sigma) \tag{3.1}$$

y is the dependent variable— percentage changes in military spending. I model the outcome using a t-distribution with degrees of freedom ν to address heavy tails and estimate ν directly. σ is analogous to the error term in a frequentist regression as it captures unexplained variation. μ, the mean of the outcome, depends on several factors.

$$\mu = \alpha + \alpha^{st} + \alpha^{yr} + \mathbf{W}_{n\times k}\gamma_{k\times 1} + \mathbf{Z}_{n\times a}\lambda_{a\times 1} \tag{3.2}$$

Percentage changes in spending are a function of an overall intercept α, state and year varying intercepts α^{st} and α^{yr} and a matrix of state-level control variables $\mathbf{W}$. The $\mathbf{Z}\lambda$ term incorporates alliance participation. $\mathbf{Z}$ is a matrix of state participation in alliances. Columns correspond to each of the a alliances in the data, and rows to state-year observations. For this sample of non-major powers, $\mathbf{Z}$ has 190 columns and 8,290 rows. If a state is not in a particular alliance, the correspon-

ding matrix cell in that alliance column is zero. If a state is part of the alliance in a given year, the matrix element contains the log of total allied military spending, which is normalized by year.[25] I use total allied spending to express the effect of alliance participation to match the theoretical emphasis on allied capability. $\mathbf{Z}$ encodes a quasi-spatial indicator of alliance participation for all a alliances in the data. States can be members of multiple treaties, so state-year observations are not neatly nested within alliances. This specification means each alliance has a unique impact on military spending, even when states participate in multiple treaties.

λ is a vector of parameters which estimate the impact of participation in specific alliances on military spending. Because the non-zero elements of $\mathbf{Z}$ are allied spending, the λ parameters capture alliance members' response to allied capability. Each alliance has a unique λ, so there are 190 alliance participation parameters. The λ parameters have a shared distribution, so I assume alliances are similar but different in how they impact military spending.

The second part of the multilevel model uses alliance characteristics to predict how alliance participation is associated with percentage changes in military spending. The alliance participation parameters are the outcome in an alliance-level regression. As a result, the impact of alliance participation on members' military spending depends on treaty characteristics, including depth. In this second-level regression:

$$\lambda_a \sim N(\theta_a, \sigma_{all}) \tag{3.3}$$

and

$$\theta_a = \alpha_{all} + \beta_1 \text{treaty depth} + \mathbf{X}_{a\times l}\beta \tag{3.4}$$

In the alliance-level regression, $\mathbf{X}$ is a matrix of the l alliance-level control variables and α_{all} is the constant. Adding σ_{all} means predictions of λ are not deterministic— the alliance level regression contains an error term. A larger σ_{all} indicates more variation in how alliance participation impacts military spending.

[25] Normalization keeps the parameters on similar scales, which is important for modeling.

The second-level regression includes treaty depth, and each β parameter modifies the impact of alliance participation on percentage changes in military spending. The βs are like marginal effects in an interaction. Treaty depth impacts military spending by modifying the consequences of alliance participation. Changing treaty depth shifts λ, which in turn affects military spending. Hypothesis 3 predicts β_1 will be negative for non-major powers.

In this model, the λ parameters express heterogeneous effects of participation in individual alliances. The β parameters estimate how alliance characteristics modify the impact of alliance participation on military spending. Again, using alliance characteristics to predict the impact of alliance participation matches my conditional argument. I now describe the sample and key variables in the analysis.

3.3.3 Sample and Key Variables

I estimate the multilevel model on a sample of non-major power states from 1816 to 2007, including state-year observations with no alliances. I identify non-major powers using a measure of major power status from the Correlates of War Project. Alliance participation data comes from the ATOP project (Leeds et al., 2002). I focus on participation in defensive and offensive treaties, because prior studies of alliances and military spending examine these treaties. The sample contains data from 8,280 state-year observations and 190 alliances.

The dependent variable is percent changes in military spending, which is calculated as:

$$\% \text{ Change Mil. Expend} = \frac{\text{Change Mil. Expend}_t}{\text{Mil. Expend}_{t-1}} \tag{3.5}$$

I used the Correlates of War Project's data on military spending to measure percentage changes in spending (Singer, 1988).The annual percentage change in spending equals that year's change in spending as a share of the previous year's military spending. Thus, annual changes are bench marked to previous spending levels. To facilitate model fitting, I apply the inverse hyperbolic sine transformation to this variable.[26] Using percentage changes in military expenditures as the depen-

[26]This transformation applies to positive, negative and zero values. It has minimal impact on values between -1 and 1, but pulls in large positive values, which range as high as 140. Inferences about treaty depth and other alliance

dent variable helps the research design. The level of military spending is not stationary for most states, especially in longer panels. Thus, using percentage changes in spending reduces the risk of spurious inferences. Benchmarking changes to prior expenditures also facilitates comparisons across states and over time.

The key independent variable is the mean latent depth of each alliance, based on the measurement model. This variable enters the model in the alliance-level regression and I expect it will have a negative coefficient. I also include a series of state and alliance-level controls.

In the state-level regression, I adjust for several correlates of alliance participation and military spending. State-level covariates include GDP growth (Bolt et al., 2018) regime type, international war (Reiter, Stam and Horowitz, 2016), civil war participation (Sarkees and Wayman, 2010), annual MIDs (Gibler, Miller and Little, 2016), rival military spending (Thompson and Dreyer, 2012) and a dummy for Cold War years. Conflict participation, alliances, and military spending are all correlated (Senese and Vasquez, 2008). I include growth in GDP instead of levels because GDP levels are non-stationary and economic growth shapes the opportunity costs of military spending (Kimball, 2010; Zielinski, Fordham and Schilde, 2017).

Other alliance level variables are correlates of treaty design and military spending, including the number of members and share of democracies in a treaty at time of formation (Chiba, Johnson and Leeds, 2015). I control for issue linkages by creating a dummy indicator of whether the alliance promises economic cooperation (Poast, 2013; Long and Leeds, 2006). As an indicator of hierarchical security relationships, I include a count of foreign policy concessions in the alliance. I also mark the presence of unconditional military support using a dummy variable I constructed using existing indicators of conditional support in the ATOP data. Because threat may drive states to form deeper alliances and affect subsequent military spending, I control for the average threat of alliance members at the time of alliance formation using the threat measure of Leeds and Savun (2007). I adjust for superpower membership— whether the United States or Soviet Union participated in a treaty during the Cold War. Two dummy indicators of wartime alliances and

characteristics are comparable with and without the transformation.

asymmetric obligations (Leeds et al., 2002) complete the alliance-level regression specification. Though I discuss these variables as controls, many of them are theoretically interesting in their own right. Having described the measure of treaty depth, multilevel model, and covariates, I now turn to the results of the analysis.

3.4 Results

This section summarizes inferences from the multilevel model. I find support for all three hypotheses. Because shallow alliances tend to increase military spending and deep alliances often decrease spending, treaty depth and the effect of alliance participation on non-major power military spending are negatively correlated. Results are based on 2,000 samples from four chains, with 1,000 warm-up iterations. To facilitate model fitting, I employed a non-centered parameterization of the varying intercepts and a sparse matrix representation of $\mathbf{Z}$. Standard convergence diagnostics indicate the chains adequately explored the posterior.[27]

Because I used Bayesian modeling to estimate the association between treaty depth and percentage changes in military spending, each coefficient has a posterior distribution— the likely parameter values conditional on the priors and observed data. There are no indicators of statistical significance. Instead, I use the 90% credible intervals of the parameters and calculate the negative posterior probability for the treaty depth coefficient to assess Hypothesis 3.[28]

Figure 3.2 summarizes the coefficient estimates from the alliance-level regression and the simulated substantive effect of greater treaty depth. The preponderance of evidence matches Hypothesis 3, as shown in the top panel of Figure 3.2. There is a 96% chance treaty depth is negatively correlated with the impact of alliance participation on percent changes in military spending for non-major powers. As treaty depth rises, the effect of alliance participation on military spending falls.

I plot the substantive effect of the alliance-level depth coefficient in the bottom panel of Figure 3.2. I assess this substantive effect by simulating the effect of changing treaty depth from the

[27] See the appendix for details on convergence and other robustness checks.

[28] I use 90% intervals because inferences about 95% intervals are sensitive to simulation variance in Bayesian analysis.

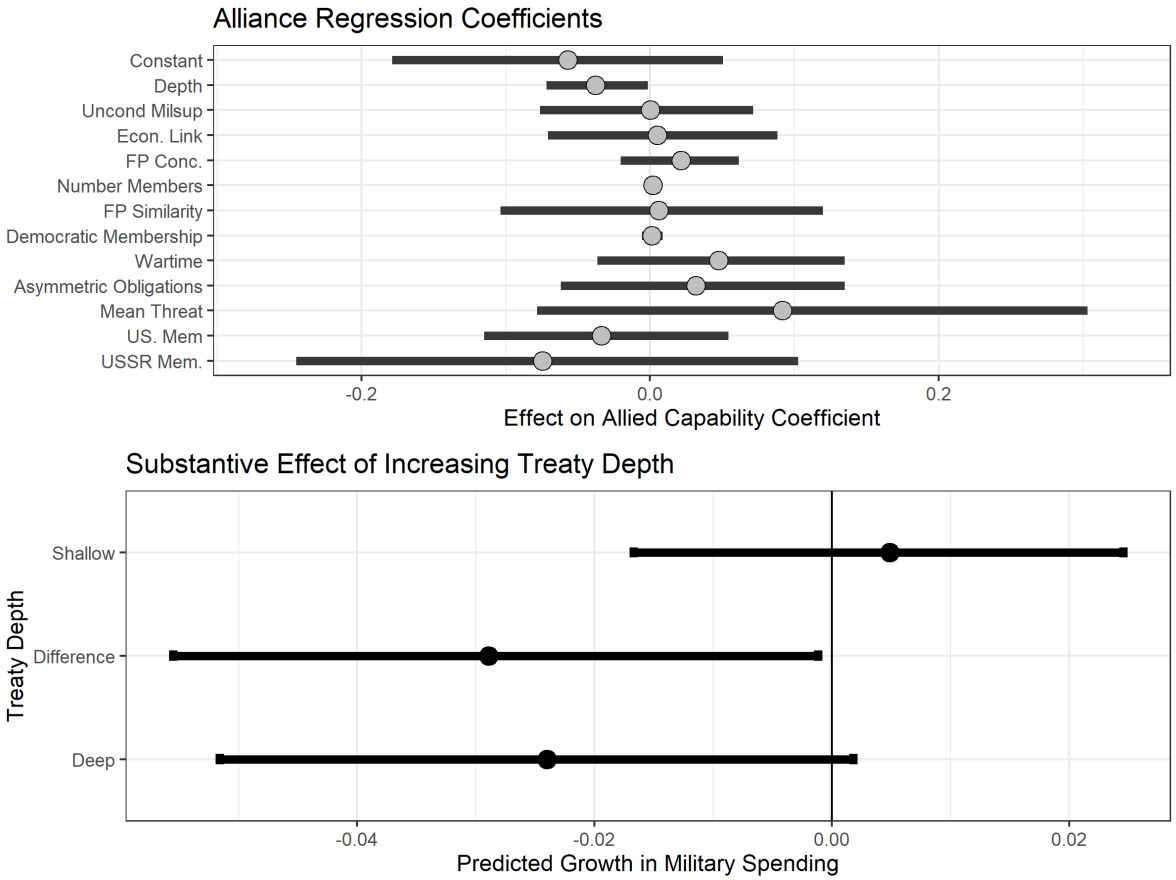

Figure 3.2: Summary of alliance-level regression results from the multilevel model of alliance participation and military spending. The top panel shows the 90% credible intervals for coefficients in the alliance-level regression. The bottom panel plots the estimated effect of participation in an alliance with average capability on growth in military spending for a deep and shallow treaty, as well as the difference between deep and shallow alliances. In both panels, points mark the posterior mean, and the bars encapsulate the width of the 90% credible interval.

minimum value of -0.8 to 1.5. Holding other alliance covariates at their modes or medians, this increase in depth reduces a hypothetical alliance parameter λ by .08 in expectation, which then lowers military spending growth. The results of the the simulation are summarized by 90% intervals for predicted growth in military spending in the hypothetical shallow alliance, predicted spending growth in the hypothetical deep alliance, and the difference between those two scenarios. Assuming the hypothetical alliance has median capability, the difference in spending growth between the shallow and deep treaty has a mean of -.03. The 90% credible interval of this predicted fall in military spending due to increasing treaty depth ranges from -0.056 to -0.001. Although the range of estimated substantive effects includes small values, there is a perceptible difference between non-major power military spending growth in deep and shallow alliances, all else equal.

To assess Hypotheses 1 and 2, I examine patterns in the alliance participation parameters across the range of treaty depth. Each λ measures the impact of treaty participation, so if treaty depth has a large influence on alliance participation, it will appear in the λ estimates. On average, participation in deep alliances should have a negative effect on members' percent changes in military spending and shallow alliances should have a positive effect. If this is true, there will also be a negative trend in the expected value of λ as treaty depth increases.

The preponderance of evidence in Figure 3.3 matches the predictions of Hypotheses 1 and 2. The top panel of Figure 3.3 plots the expected value of the alliance participation parameter across the range of treaty depth. As expected, shallow treaties often have positive λ values,[29] which corresponds to Hypothesis 1. Most of the deepest treaties have a negative λ, which matches Hypothesis 2. Because other treaty characteristics and chance also influence the λ estimates, there is substantial variation in how alliance participation impacts non-major power military spending.

I then used the λ posteriors to estimate the impact of alliance participation on state-year military spending growth. To assess the annual impact of alliances on their members, I multiplied the alliance membership matrix $\mathbf{Z}$ by the λ parameters to generate 9,124 non-zero state-alliance-year

[29] All the negative λ estimates in alliances with minimal depth are treaties between the Soviet Union and Eastern European states during the Cold War.

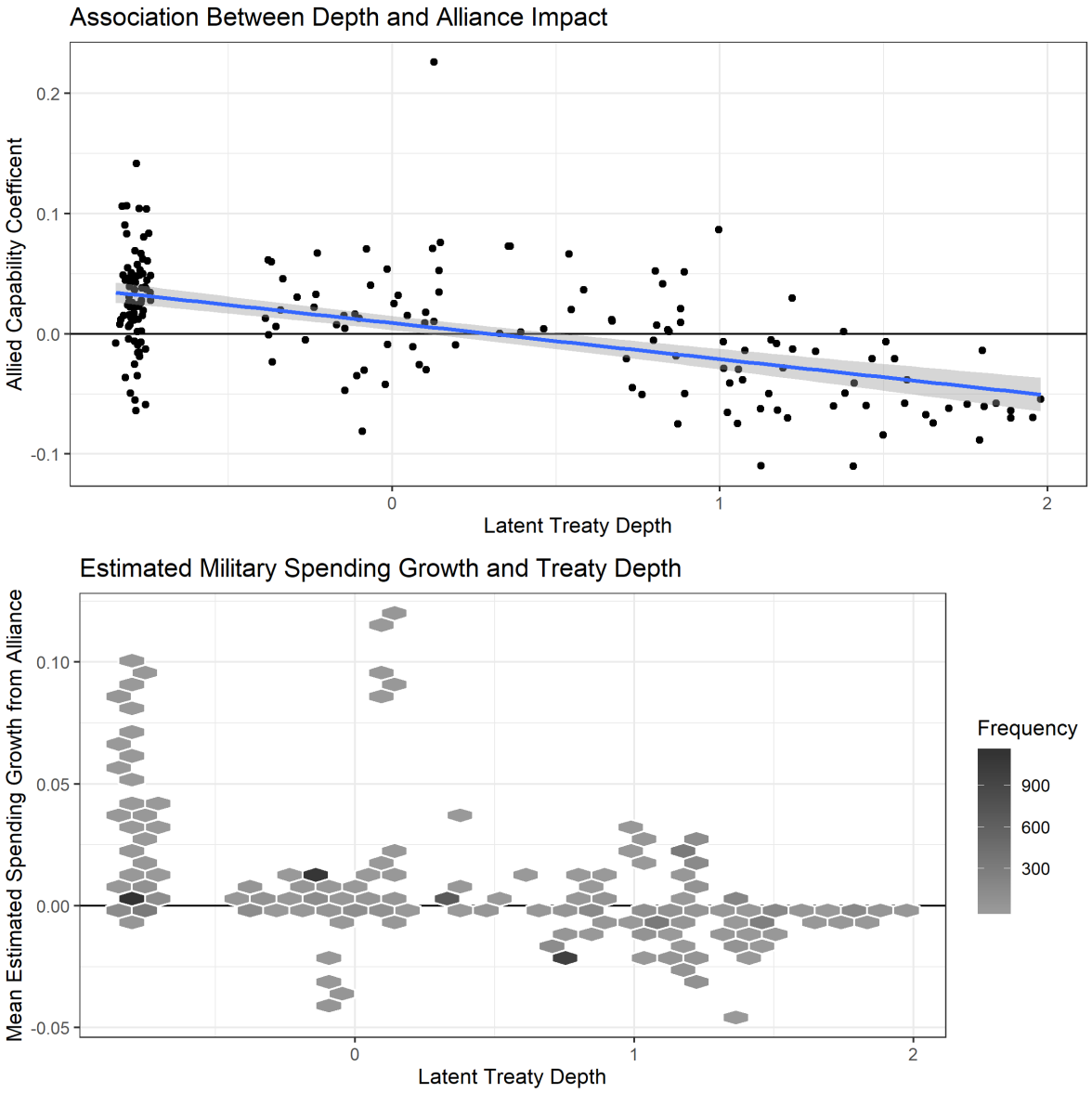

Figure 3.3: Summary of the predicted effect of alliance participation on growth in military spending across the observed values of treaty depth for 190 alliances from 1816 to 2007. The top panel plots the mean of each λ parameter by treaty depth. The bottom panel plots 9,128 state-alliance-year estimates of how participation in individual alliances affects growth in military spending, which are the product of the λ for the alliance and allied capability value for that state-year observation. Darker shading indicates more data points in the hexagon.

predictions.[30] The scatter plot in the bottom panel of Figure 3.3 shows the distribution of predicted changes in military spending from alliance participation. Each point marks the mean estimated effect of an alliance on growth in military spending for that state-year. To avoid overplotting the 9,124 point estimates, I combined them into hexagons. Darker hexagons mark areas with more points. As Hypothesis 1 predicts, participation in shallow alliances regularly increases military spending. Many alliances with shallow depth have little effect on military spending, however. As treaty depth increases, alliance participation is more likely to reduce military spending growth.[31] Some of the deepest alliances have a negligible effect on military spending despite negative λ values due to limited capability.

In summary, I find that treaty depth modifies the impact of alliance participation on military spending. Participating in deep alliances often reduces military spending, while being part of a shallow alliance often increases spending. I now discuss the implications of the argument and results.

3.5 Discussion and Conclusion

This paper addresses a longstanding debate over whether alliance participation increases or decreases military spending. Claims alliance participation only increases or decreases military spending are incomplete. My argument shows how treaty depth modifies the impact of alliance participation on military spending, which builds on other conditional arguments (DiGiuseppe and Poast, 2016). I show that whether alliance participation increases or decreases military spending depends on treaty depth. Compared to no alliance at all, joining a shallow treaty usually increases military expenditures, while participation in a deep alliance often lowers defense spending.

There are two other noteworthy findings. First, many alliances increase non-major power military spending, which cuts against expectations that non-major powers are inveterate free-riders. To gain security from alliance participation, non-major powers sometimes need to increase their defense budget (Horowitz, Poast and Stam, 2017). Second, the military spending growth predictions

[30] These estimates hold all state-level variables constant.

[31] Wartime alliances are the main exception to this trend.

in Figure 3.3 suggest that many alliances have little effect on military spending, which is puzzling given widespread expectations that alliances and military spending are related. These negligible effects could reflect offsetting effects from different parts of alliance treaty design and membership, or concerns about the credibility of many alliance treaties.

Given my departure from previous research designs, how should we compare these results to prior evidence on alliance participation and military spending? Connecting my results with earlier evidence requires renewed attention to specific and general research designs. Recall that general studies compare states in an alliance to those without one in a global sample and specific studies estimate responses to allied military spending in a few alliances. The results encompass specific and general research designs by using allied capability to measure alliance participation. Each alliance participation parameter includes the effect of joining an alliance and changes in allied capability. I then use the alliance-level regression to understand how the impact of alliance participation varies with treaty design and membership.

Although my research design advances the debate on alliance participation and military spending, it has two limitations. First, my findings only address formal treaty depth. The measure of treaty depth only includes formal promises, in part because informal depth is harder to observe. As a result, my test of alliance depth may be conservative— it does not capture phenomena that should have a similar effect. It could instead overstate the findings if formal depth is not implemented, however. Strategic alliance design is the second possible weakness of the test. Domestic politics can affect alliance obligations, for example (Davis, 2004; Chiba, Johnson and Leeds, 2015). To address this issue, I controlled for correlates of alliance participation and treaty depth at each level of the model, with a particular focus on factors like democracy, alliance size, external threat, and other sources of credibility. At the state level, I adjusted for threat, economic growth, and regime type, all of which are possible correlates of treaty depth and growth in military spending. Even with this effort, selection into different alliances could still produce unobserved differences between alliances I did not adjust for.

Despite these limitations, the argument and results generate valuable insights about alliance

participation and military spending. I explain when alliance participation is associated with increases or decreases in military spending among non-major powers, which addresses a debate between contradictory views of alliances. I provide evidence that how alliance participation impacts military spending depends on state capability and alliance treaty depth using a new measure of alliance treaty depth and a multilevel model.

There are several implications for scholarship. First, my argument and findings reinforce the importance of accounting for heterogeneity among alliances and institutional design more generally. Second, my research design could apply to other international institutions where institutional design shapes the consequences of participation. Besides these general implications, the results raise interesting questions for future research. For example, why do many alliances have a negligible effect on non-major power military spending? Because arms and alliances both provide security, existing arguments expect they are connected, but I find that many alliances have little substantive effect. The domestic economic and political consequences of alliance participation also remain relatively unexplored. Research on this topic could explore how changes in allied military spending affect public support for alliance participation and the economic consequences of changes in military spending.

Besides their scholarly value, the argument and evidence help inform policy debates about military spending. My argument claims that reassurance from deep alliances leads to lower defense spending. States can use deep cooperation to increase alliance credibility, but allied military spending may fall in response. Therefore, there is a tradeoff between treaty credibility and allied military spending.

The United States is currently wrestling with the credibility-military spending tradeoff. Washington has often decried allies who provide too little for their own defense (Lanoszka, 2015). But allies are able to maintain low military spending partly because the United States makes deep commitments. Reducing the depth of US alliances could generate credibility problems, however. Low allied defense spending may be the price of credible commitments. Therefore, this paper is not an unconditional call to reduce the depth of US alliances. Adjusting existing treaties may be more

difficult than designing new alliances and could have other ramifications. The full consequences of shifting treaty depth require additional scrutiny.

4. DEMOCRACY AND ALLIANCE TREATY DEPTH

Why do states make deep alliance treaties? While some alliances only offer military support, others add further commitments of peacetime defense cooperation and policy coordination. Deep alliances include promises like basing rights, international organizations, policy coordination, and integrated military command. I argue that states use treaty depth to increase the credibility of their alliance commitments while managing the risk of entrapment.

The sources of treaty depth are worth studying because states frequently employ treaty depth. Moreover, treaty depth changes alliance politics. Half of all ATOP alliances with offensive or defense obligations have some treaty depth. Credibility from deep alliances encourages non-major power members to reduce military spending. Thus, treaty depth affects alliance politics by shaping treaty credibility and the distribution of military spending among members.

Despite the consequences of alliance treaty depth, we have little idea when states add depth to their alliances.[1] In this paper, I explain when states make deep alliance treaties. I start with the premise that depth is one of several ways that states can increase the credibility of their alliance commitments, and argue that domestic political institutions lead states to prefer treaty depth to other sources of credibility. Treaty depth and unconditional military support are two costly ways states can increase the credibility of their alliance commitments. Democracies prefer to reassure allies with deep alliance treaties because depth increases alliance credibility with less exposure to entrapment or domestic audience costs. Unconditional military support exposes leaders to entrapment or the audience costs of treaty violation, but limited commitments of military intervention are harder to violate (Mattes, 2012*b*; Chiba, Johnson and Leeds, 2015). Therefore, I expect that democracies will often form deep alliances with conditional promises of military support.

I test the argument with a statistical analysis of offensive and defensive alliances from 1816 to 2007 and an illustrative case study of the North Atlantic Treaty Organization (NATO). I estimate a series of statistical models, including several bivariate models that adjust for unobservable cor-

[1] Mattes (2012*b*) examines the causes of military institutionalization.

relations between treaty depth and unconditional military support (Braumoeller et al., 2018). The case study shows that democratic alliance leaders fear entrapment, but are willing to use treaty depth to reassure allies (Seawright and Gerring, 2008; Seawright, 2016). I find consistent evidence that alliances with democratic leaders have higher treaty depth. Evidence that democratic alliance leadership decreases the probability of unconditional military support is less consistent, however.

This paper contributes to three strands of scholarly inquiry. First, they add to a growing literature on alliance treaty design. Existing scholarship examines parts of treaty design in isolation (Benson, 2012; Mattes, 2012*b*; Chiba, Johnson and Leeds, 2015), but I show how different sources of credibility in alliance treaty design are connected. My argument considers how audience costs shape preferences for different types of reassurance, and some empirical models account for the relationship between different sources of alliance credibility.

Second, this paper adds to knowledge of how domestic politics affect alliance politics. Scholars have long acknowledged that democracy and alliances are connected (Lai and Reiter, 2000; Gibler and Wolford, 2006; Mattes, 2012*b*; Warren, 2016; McManus and Yarhi-Milo, 2017). Existing scholarship suggests that democracies prefer limited commitments (Mattes, 2012*b*; Chiba, Johnson and Leeds, 2015; Fjelstul and Reiter, 2019) because they are more likely to make conditional promises of military support. Chiba, Johnson and Leeds (2015) write that "domestic costs can make democratic states wary of engaging in agreements requiring broad and/or deep cooperation." My argument suggests that even if democracies screen the scope of their commitments carefully, they form deeper alliances on other dimensions. As a result, the net effect of the connection between democracy, alliance design and treaty credibility requires further research. Scholars often claim that audience costs make democratic commitments more credible (Gaubatz, 1996; Leeds, Mattes and Vogel, 2009; DiGiuseppe and Poast, 2016), but this relationship is disputed (Gartzke and Gleditsch, 2004; Downes and Sechser, 2012). Third, I contribute to the rich literature on domestic politics and international cooperation, e.g. (Downs and Rocke, 1995; Fearon, 1998*b*; Leeds, 1999; Mattes and Rodríguez, 2014). If democracies reassure partners with deep alliance commitments, audience costs sometimes push democracies to undertake international commitments with

less electoral salience.

Despite the importance of alliance treaty depth, the process of alliance treaty negotiation and design is understudied (Poast, 2019*a*), and there is even less research on treaty depth. In the nascent alliance treaty design literature, most scholarship examines conditions on military support. Benson (2012) shows that foreign policy disagreements and revisionist protege states increase the likelihood of limited military support and Kim (2011) claims that states use conditional obligations to escape entrapment. Mattes (2012*b*) finds that joint democracy increases the probability of conditional alliance obligations, as well as little evidence that alliance symmetry and history of violation affect conditionality. Chiba, Johnson and Leeds (2015) add to to existing work on limited obligations with evidence that democracies are more likely to form alliances with conditional military support or consultation.

Two studies that examine concepts that are similar to treaty depth, but both treat depth and conditions on military support as independent sources of credibility. First, Mattes (2012*b*) argues that members of symmetric bilateral alliances where one partner has history of violation will be more likely to use conditions on military support, issue linkages, and military institutionalization to increase treaty reliability. Her measure of military institutionalization (Leeds and Anac, 2005) is close to my conceptualization of treaty depth, and Mattes finds that symmetric alliances where one partner has a history of violation have higher institutionalization. This paper makes an important contribution, but it does not differentiate between the costs and benefits of different sources of alliance credibility, and the military institutionalization measure understates the extent of variation in treaty depth. Second, Benson and Clinton (2016) find that foreign policy agreement, major power involvement and treaty scope are all positively correlated with depth, while checking the validity of a latent measure of how costly alliance obligations are in general. Benson and Clinton's latent measure of depth includes secrecy and issue linkages, so it captures a broader concept than deep military cooperation. Neither study provides an explanation for why states might prefer depth to other sources of alliance credibility.

Therefore, we still do not understand why states add treaty depth to their alliances, especially

in relation to other sources of alliance credibility. To explain this, I compare treaty depth to unconditional military support. Both treaty depth and unconditional military support increase alliance credibility, but they have different costs. Sensitivity to those costs especially entrapment, audience costs, and foreign entanglement depends on domestic political institutions.

The paper proceeds as follows. In the next section, I lay out the argument and hypothesis. Then I describe the data and research design. The final sections discuss the results and implications.

4.1 Argument

In this argument, I first establish a definition of treaty depth. After that, I offer a general explanation of the costs and benefits of treaty depth, relative to unconditional military support. Based on the general framework, I then detail why democracies often increase treaty depth, but are less likely to offer unconditional military support.

Alliance depth is the extent of defense cooperation a treaty formalizes. Deep alliances require additional military policy coordination and cooperation. While shallow alliances stipulate more arms-length ties between members, deep treaties lead to closer cooperation through intermediate commitments that fall between treaty formation and military intervention. Defense cooperation in a deep alliance takes many forms. Allies can promise an integrated military command, military aid, a common defense policy, basing rights, international organizations, specific capability contributions or companion military agreements.

Treaty depth helps states address problems of opportunism with commitments of military support. Alliances are self-enforcing contracts or institutions (Leeds et al., 2002; Morrow, 2000). Given external threats in an anarchic international system, states form alliances to aggregate military capability and secure their foreign policy interests (Altfield, 1984; Smith, 1995; Snyder, 1997; Fordham and Poast, 2014). The treaties that formalize promises of military support take many forms (Leeds, Long and Mitchell, 2000; Leeds et al., 2002; Benson, 2012; Benson and Clinton, 2016). Treaty design shapes the costs and benefits of treaty participation and addresses potential opportunistic behavior. Beyond the benefit of potential military support, alliances also clarify international alignments (Snyder, 1990) and support economic ties (Gowa, 1995; Li, 2003; Long,

2003; Fordham, 2010; Wolford and Kim, 2017). The costs of alliances include lost foreign policy autonomy (Altfield, 1984; Morrow, 2000; Johnson, 2015), and the potential consequences of opportunistic behavior. Opportunism in alliances includes abandonment, or the failure of alliance members to honor their commitments (Leeds, 2003*a*; Berkemeier and Fuhrmann, 2018), entrapment in unwanted conflicts (Snyder, 1984), and free-riding (Morrow, 2000).

To form an alliance, states must have sufficient overlap in foreign policy interests (Morrow, 1991; Smith, 1995; Fordham and Poast, 2014), especially their proposed war plans (Poast, 2019*a*). By making a formal alliance treaty, members attempt to increase the credibility of military interventions (Morrow, 2000). Alliance members and other states can then use the costs of the alliance commitment to assess treaty reliability. While alliance formation alone adds some credibility, treaty depth or promises of unconditional military support further increase alliance credibility. Depth and unconditional military support are both costly commitments that alliance members and other states can observe.

Greater alliance credibility can increase the risk of entrapment in unwanted conflicts, however. Whether an alliance imposes conditions on military support clearly shows this tradeoff. Conditional alliances limit promises of intervention to particular regions, conflicts, or instances of non-provocation (Leeds, Long and Mitchell, 2000). When alliance members fear entrapment in unwanted conflicts because their allies have divergent interests, they constrain military support to specific circumstances (Kim, 2011; Benson, 2012).[2] Conversely, offering unconditional military support indicates high shared foreign policy interests and less fear of entrapment. Attaching no conditions to a promised intervention means alliance members hazard the reputational (Gibler, 2008; Crescenzi et al., 2012) and audience (Fearon, 1997) costs of treaty violation from many potential conflicts. Accepting these potential costs implies that fighting with allies in many circumstances is acceptable. Therefore unconditional alliances are a key source of credibility, because they are a costly signal of substantial foreign policy agreement.

When states form alliances with unconditional promises of military support, they have few

[2]Such deliberate design of alliances means clear instances of entrapment are rare (Kim, 2011; Beckley, 2015).

entrapment concerns, but alliance members face a time-inconsistency problem (Leeds and Savun, 2007). Although conditions on military support are fixed, foreign policy interests change. Alliance members may start with substantial foreign policy agreement, only for their interests to diverge. If interests change and the alliance is invoked, states must bear either entrapment or the audience costs of treaty violation. Inasmuch as audience costs are less costly than fighting, states will often choose to violate the treaty and bear the audience costs. But states with high audience costs will be more sensitive to entrapment, because alliance violation is more costly. Therefore, states with high audience costs will be less likely to offer unconditional military support (Chiba, Johnson and Leeds, 2015).

Unconditional military support is not the only source of alliance credibility, however. Treaty depth offers a way for states with audience cost concerns to reassure allies with less risk of entrapment. Depth shapes the perceived reliability of an alliance by providing opportunities for states to fulfill treaty obligations in peacetime (Morrow, 1994). When states implement deep obligations, it provides a sunk cost signal of commitment. Observing that alliance members adhere to peacetime promises suggests that they will also honor promises of military support.

How does depth limit the risk of entrapment? Failure to implement deep alliance provisions provides an observable signal of foreign policy divergence short of war. Furthermore, policy coordination in a deep alliance gives states more influence over allied actions and policies. Bases, joint organization and policy coordination are all potential checks on entrapment. Coordinating policies and plans over time allows states to ensure that their joint war plans still line up, and that neither member faces a prohibitive risk of abandonment or entrapment.

Close cooperation through treaty depth addresses the reliability and entrapment dilemma, but it has other costs. Besides opportunism, lost foreign policy autonomy is the other major cost of alliance participation. Deep alliances reduce foreign policy autonomy more than other alliances. Cooperating and coordinating policy with allies reduces the ability of states to make unilateral decisions. There are also practical hurdles to unwinding foreign bases, international institutions and integrated military commands. In summary, depth adds credibility with less risk of entrapment,

but entails substantial foreign entanglement.

Though depth and unconditional military support both increase alliance credibility, they do so in different ways. Conditions on military support do not change without a treaty renegotiation, and the costs of fighting are hypothetical unless the alliance is invoked. Highly conditional alliances are less costly in expectation, which reduces their credibility but guards against entrapment. Conversely, the sunk costs of depth can be observed without invoking promises of military support and address time-inconsistency problems. Depth and cooperation with allies increases credibility with less risk of entrapment, but leads to greater foreign entanglement.

Whether states are more concerned with entrapment or foreign entanglement therefore shapes how they reassure their alliances. When states fear entrapment and do not want to face the audience costs of treaty violation, they will form deep alliances with conditional promises of military support.[3] Therefore, states with high audience costs for violating promises of military support and relatively low concern with foreign entanglement in the leader's key foreign policy audiences will use treaty depth to increase the credibility of their alliances. Domestic political regime type shapes leaders' foreign policy audiences and the balance of these concerns. Due to high audience costs of military support and low public concern with foreign entanglement short of military intervention, I expect that alliances with democratic leaders will be more likely to have limited military support and high depth.

4.1.1 Democratic Alliance Membership and Treaty Design

Democracies use treaty depth to reassure partners because it increases the perceived reliability of their alliances while shielding them from domestic audience costs. Violating international promises can have audience costs for democratic leaders. Democratic leaders fear that if they violate international commitments, public disapproval will lead to their removal from office. Audience costs do not apply to all international commitments, however. Two characteristics of audience costs in democracies encourage them to use treaty depth in alliance commitments. First, audience

[3]Prospective alliance members could also make an arms-length alliance commitment with neither depth nor unconditional military support, or use treaty depth to address time-inconsistency problems with unconditional military support.

costs increase as crises escalate (Tomz, 2007). Second, the public in democracies lacks substantial foreign policy information, so they are less concerned about foreign entanglements short of war and treaty depth has limited salience in domestic politics. While a potential military intervention is costly and highly public, peacetime alliance cooperation is not. Democratic voters are unlikely to have much information or strong preferences about defense cooperation with allies. Although foreign policy elites may dispute changes in commitment to a deep alliance, such dissent is unlikely to translate into meaningful public opposition and electoral concerns.

Unlike in domestic politics, differences in treaty depth and the implementation of alliance promises are salient in international politics. Allied states and potential adversaries can gather useful information from treaty depth. By including peacetime costs in a deep treaty, alliance members signal alliance reliability. Implementing costly promises of military aid, bases, or policy coordination indicates commitment. This increases allied confidence that democracies will honor their treaty obligations. Therefore, democracies can use treaty depth to signal international commitment with less exposure to domestic audience costs.

The way democracies use treaty depth in alliances is analogous to their reliance on non-tariff barriers in trade policy. Kono (2006) argues that because non-tariff barriers are more complex, voters lack sufficient information about their impact on consumer prices. Tariffs, on the other hand, translate directly into prices in ways that are easy to understand. The complexity of non-tariff barriers makes them less vulnerable to electoral attack, so democracies engage in "optimal obfuscation" and substitute non-tariff barriers for tariffs. In the same way, unconditional promises of military support and treaty depth both affect international relations, but the former is salient for voters and the latter is not. Unconditional military support is a straightforward tool for electoral attack as it is east to explain to voters— treaty depth is not. Therefore, democratic leaders can use treaty depth to manage international relations with fewer domestic political consequences.

Open political competition for leadership, contested elections and legal constraints on executive action are three possible sources of audience costs in democracies. Open recruitment of leaders through elections increases the chance a leader is removed from office. Open political competition

allows opposition groups to hold leaders to account for foreign policy shortcomings. For leaders to fear replacement, there must be a group with some chance of organizing and throwing them out of office. Democratic leaders also face legal constraints on their foreign policy actions, and violating these restraints could generate audience costs. I examine the relative weight of these three mechanisms in the empirical analysis.

This argument uses the limited domestic audience costs and reassurance benefits of treaty depth to explain why democracies often form deep alliances. How does this translate to autocracies? Some autocratic leaders, especially in single-party states, also face high audience costs for backing down in military conflicts due to scrutiny from domestic elites (Weeks, 2014). Domestic elites in single party states are a different audience than the public in democracies, however. Party elites are more informed about foreign policy than the public in democracies, so they can also impose costs on leaders that violate promises of treaty depth. In general, no autocratic audience has the combination of limited foreign policy information and high audience costs of military intervention as democracies. Single-party and military regime leaders face a highly informed domestic elite, and personalist leaders have few meaningful foreign policy audiences. Therefore, assuming that all autocracies are equivalent, relative to democracies is sufficient for this paper.[4]

Of course, democracies may not get everything they want in alliance negotiations. Following (Mattes, 2012*b*), I expect that more capable states have greater influence on alliance negotiations, because their partners lose out on more foreign policy benefits if they are out of the alliance. The most capable state is often the alliance "leader," which increases the weight of their preferences and concerns. Therefore, to understand how democracy shapes alliance treaty design, I conceptualize democratic influence in terms of the political regime type of the most capable alliance member.

Due to the limited audience costs of treaty depth, democracies will often design deep alliance treaties to increase the credibility of their alliance commitments. As the democracy of the most capable alliance member at the time of treaty formation increases, treaty depth should increase.

[4]Examining heterogeneity among autocracies in alliance treaty design is an interesting subject for future research, however.

TREATY DEPTH HYPOTHESIS: AS THE AVERAGE DEMOCRACY OF THE MOST CAPABLE ALLIANCE MEMBER AT THE TIME OF FORMATION INCREASES, ALLIANCE TREATY DEPTH WILL INCREASE.

I also expect that democratic alliance membership will reduce the probability of unconditional military support, because backing out of a promised military intervention generates substantial audience costs. Analyzing unconditional military support compares treaty depth with another source of credibility. The second claim is based on existing arguments and findings. Mattes (2012*b*) and Chiba, Johnson and Leeds (2015) both show that democracies are more likely to design conditional alliances. They attribute this finding to higher audience costs of violating international commitments in democracies. Because democratic leaders face substantial audience costs from violating their international commitments, leaders design more limited commitments that are easier to fulfill. On the other hand, autocracies may be more willing to promise unconditional military support, because backing out of unwanted conflicts has lower audience costs. Based on this logic, increasing the democracy of the most capable member when the alliance formed should reduce the probability of unconditional military support.

UNCONDITIONAL MILITARY SUPPORT HYPOTHESIS: AS THE AVERAGE DEMOCRACY OF THE MOST CAPABLE ALLIANCE MEMBER AT THE TIME OF FORMATION INCREASES, THE PROBABILITY THAT THE ALLIANCE OFFERS UNCONDITIONAL MILITARY SUPPORT WILL DECREASE.

Depth and conditional obligations in democratic alliances are complementary, as depth substitutes for unconditional military support as a source of credibility. Limiting alliance commitments through conditional military support reduces audience costs because it is easier for democratic leaders to backtrack on military support. Leaders can claim that the conditions for intervention were not met, or that new information obviates the alliance commitment (Levendusky and Horowitz, 2012). Because allied states understand these limits, conditional alliances will increase reliability concerns. This may undermine the credibility of the alliance, as unreliable pacts invite challengers

(Smith, 1995). But even if conditional military support reduces the credibility of democratic alliances, treaty depth strengthens the alliance. Thus, depth is an important source of reassurance for the allies of democracies, who might otherwise have reliability concerns from conditional military support.

In previous research, scholars often use the finding that democracies prefer conditional alliances to claim that democracies make limited alliance commitments. But if democracies frequently add depth to their alliance treaties, democratic alliances are limited on one dimension and deep on others. Furthermore, deep alliances may produce closer security ties than unconditional military support.

There is an important caveat to this argument— I am interested in institutional design, not implementation. Alliance treaty depth is not always implemented fully, as treaty aspirations are not fully realized, or work poorly. To give one example, several deep Arab alliances never realized their full intention due to internal political divisions. If states are to use treaty depth to increase treaty credibility, they must implement some of their alliance promises, however.

I expect that more democratic alliance membership will increase treaty depth, but reduce the likelihood of unconditional military support. This occurs because democracies prefer to use treaty depth to reassure their partners. In the next section, I describe how I test this claim about the association between democratic alliance leadership and the two sources of alliance credibility.

4.2 Research Design

I expect that democracy among alliance members will increase treaty depth and decrease the probability of unconditional military support. Furthermore, the processes behind conditions on military support and treaty depth may be related, as both are sources of credibility. I examine these claims with a series of statistical models and an illustrative case study of NATO. I first describe the key variables in the analysis, then provide more detail on the estimation strategy.

To examine my predictions that democracies tend to produce conditional alliances with substantial depth, I employ data from the Alliance Treaty Obligations and Provisions Dataset (Leeds et al., 2002). The sample includes 289 alliances with either offensive or defensive obligations,

which is the set of treaties with military support.[5] I measured treaty depth using a semiparametric mixed factor analysis of eight ATOP variables (Murray et al., 2013).[6] The depth measure is a weighted combination of ATOP's defense policy coordination, military aid, integrated military command, formal organization, companion military agreement, specific contribution, and bases variables. Each of these individual indicators increases alliance treaty depth, but defense policy coordination and an integrated command have the largest positive association, as shown in the top panel of Figure 4.1.

Based on these factor loadings, the measurement model predicts the likely value of treaty depth. The distribution of depth is summarized by the bottom panel of Figure 4.1. There is substantial variation in alliance treaty depth. Around half of all formal alliance treaties have an least some depth, and there is wide variation in how much depth is present. I measure treaty depth using the posterior mean of the latent depth posterior for each alliance. This summarizes the central tendency of latent treaty depth, and results are robust to accounting for uncertainty in the latent measure.

The other outcome variable is a dummy indicator of unconditional military support. Using ATOP's information on whether defensive or offensive promises are conditional on specific locations, adversaries, or non-provocation, I set this variable equal to one if the treaty placed no conditions on military support. 123 of 289 alliances in the data offer unconditional military support.

The key independent variable is the democracy score of the most capable alliance members when the treaty formed. I use the POLITY measure of political institutions to measure democracy, and code the alliance leader as the state with the largest CINC score (Singer, 1988). This measure captures the sensitivity of the leading state in the alliance to audience costs and entrapment. It also emphasizes the influence of the most capable alliance member.[7]

A common alternative measure of allied democracy is a dummy variable which is equal to one

[5] Results are robust to adjusting for non-random selection into alliances. See the appendix for details.

[6] **https://github.com/joshuaalley/arms-allies/blob/master/manuscript/arms-allies-paper.pdf** contains more details on the measure.

[7] I find similar results with a model that uses the proportion of democracies as the key independent variable, which I report in the appendix. The proportion of democracies captures the prevalence of democracies, and it has a strong positive correlation with the democracy of the most capable member.

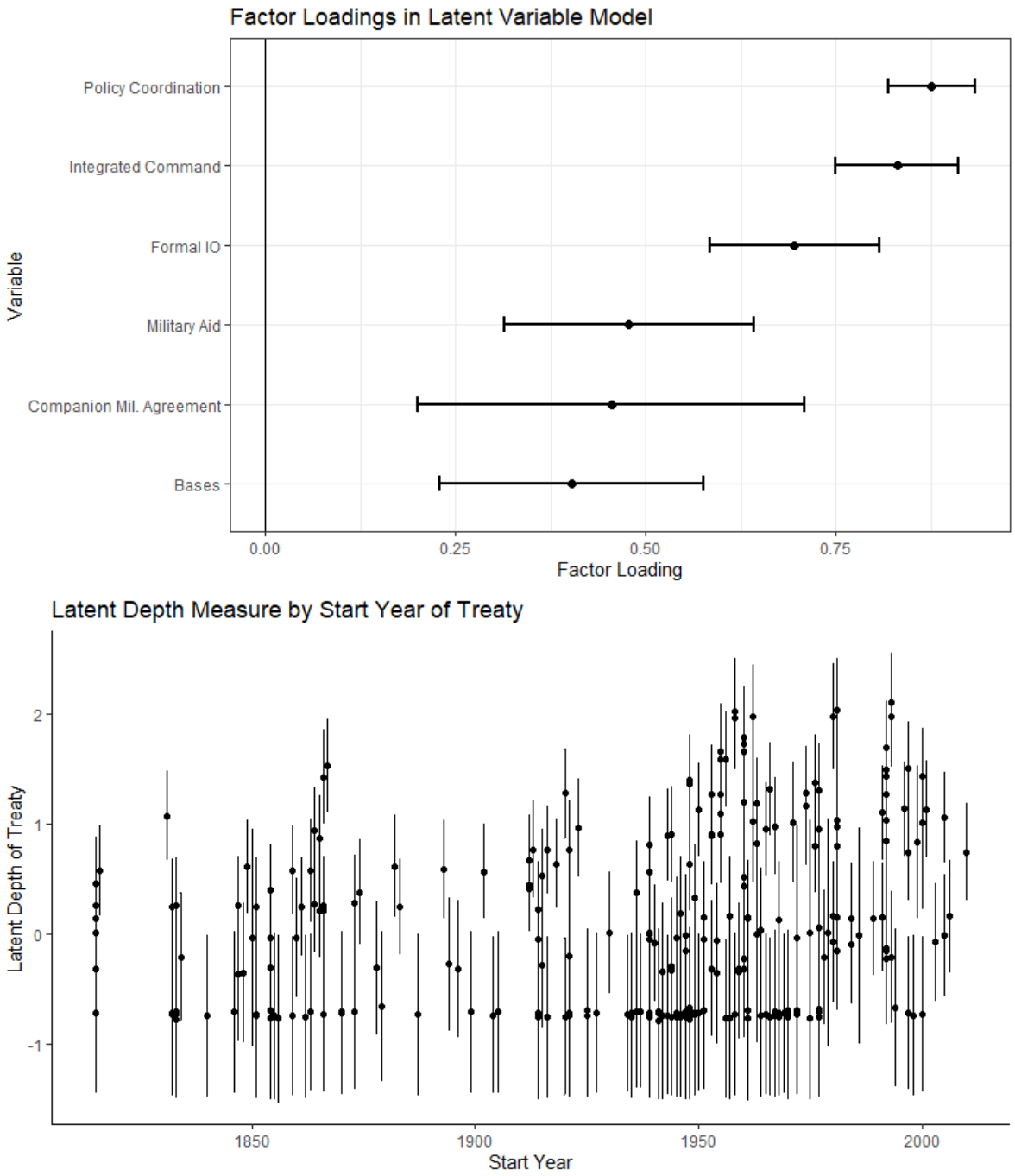

Figure 4.1: Factor Loadings and posterior distributions of latent alliance treaty depth measure.

if both alliance members have a polity score greater than 5.[8] I prefer the leading alliance member democracy score to this dummy for two reasons. First, it translates better to multilateral alliances. Moreover, though alliance leader democracy and joint democracy are correlated, joint democracy is a slightly different concept. I expect that domestic regimes affect alliances between democracies and non-democracies, even if such alliances are unusual (Leeds, 1999).

After estimating the association between overall POLITY scores and alliance treaty design, I break the democracy measure into three components to identify the effects of differnt institutional features of democracy. The POLITY measure combines executive recruitment, political competition and executive restraints. To examine each separately, I created three dummy variables, one for each concept. The dummies are equal to one if POLITY codes the most capable state as having competitive elections for leadership, open political competition and executive parity/subordination, respectively. After presenting the aggregate democracy results, I show inferences from a model with these three dummies as the key independent variables.

4.2.1 Estimation Strategy

I use several statistical models to examine how political regimes affect treaty depth. I start with separate models of treaty depth and unconditional military support. But because common unobserved factors may affect depth and conditionality, I also specify a bivariate model with correlated errors as a robustness check. Without modeling the correlation between depth and unconditional military support, univariate models may produce biased estimates. This approach is analogous to a bivariate probit model, but it is not fully recursive, because I do not use depth or unconditional military support as endogenous predictors of the other factor.[9]

To predict unconditional military support, I fit a binomial model with probit link function. The alliance leader democracy measure is the key independent variable, and I control for a range of other factors that are likely correlates of unconditional military support and allied democracy. Key controls include dummy indicators of asymmetric alliances between non-major and major powers

[8] Only 25 of the 285 alliances in my data clear this joint democracy threshold.

[9] A fully recursive model requires instruments for identification.

and symmetric alliances between major powers (Mattes, 2012*b*)[10] as well as the average threat among alliance members at the time of treaty formation (Leeds and Savun, 2007). I also control for foreign policy similarity using the minimum value of Cohen's κ in the alliance (Häge, 2011). I draw on the ATOP data (Leeds et al., 2002), to adjust for for asymmetric treaty obligations, the number of alliance members, whether any alliance members were at war and the year of treaty formation. To capture the role of issue linkages in facilitating alliance agreements (Poast, 2012, 2013), I include a dummy indicator of whether the alliance made any economic commitments. Last, I include a count of foreign policy concessions in the treaty, because concessions can facilitate agreement in alliance negotiations (Johnson, 2015).

The model of treaty depth controls for the same set of variables as the model of unconditional support. Modeling depth is more complicated because the latent measure is skewed. To facilitate model fitting, I transformed latent depth to range between zero and one and modeled it with a beta distribution.[11] The flexibility of the beta distribution helps predict mean latent depth.[12]

To start I fit the depth and unconditional support models separately. Then, as a robustness check, I use a generalized joint regression model (GJRM) (Braumoeller et al., 2018) to fit the models of unconditional support and treaty depth simultaneously. GJRM uses copulas to model correlations in the error terms of multiple equation models, which makes it more flexible than parametric models and facilitates causal inference. Adjusting for unobserved correlations between depth and unconditional military support ensures accurate inferences about democracy and other covariates. Copulas are distributions over functions, and relax potentially problematic assumptions about the shape of the correlation in the error terms. I fit models with all copulas, and selected the best-fitting model using AIC, conditional on that estimator having converged.[13] The T copula

[10]This leaves symmetric alliances between major powers as the base category

[11]I make similar inferences with a robust regression estimator- see the appendix for details. I also considered log-logistic, Dagum and inverse Gaussian distributions for the outcome, but AIC and residuals showed that the beta model fit best.

[12]The beta distribution also facilitates fitting models that account for uncertainty in the latent measure, which I include in the appendix.

[13]GJRM is estimated with maximum likelihood, and diagnostics for the gradient as well as the information matrix suggest that the models converged.

provides the best model fit.[14]

In general, the research design gradually increases in complexity. I start with descriptive statistics. Then I fit separate models of treaty depth and unconditional military support, followed by a joint model. Finally, I use a joint model to estimate how elections, open political competition and executive constraints affect alliance treaty design. The next section summarizes the results.

4.3 Results

My findings are partially consistent with the claim that increasing democracy in an alliance leads to treaties with conditional support and greater depth. I find consistent evidence that democracy in the alliance leader increases treaty depth, but weaker evidence about democracy and conditional obligations. These results reflect competing effects of different democratic institutions. When I disaggregate democracy into elections, open political competition and executive constraints, I find that competitive elections increase depth and decrease the probability of unconditional military support. I fund that executive constraints increase the probability of unconditional military support and offset the effect of elections, however.

Descriptive statistics are consistent with my hypotheses. On average, unconditional alliance leaders have lower polity scores. The average alliance leader polity score among alliances with unconditional military support is -2.55. The average leader polity score in alliances with conditional obligations is -.24.[15] There is also a modest positive correlation between the alliance leader's democracy at the time of formation and treaty depth.

Figure 4.2 shows how the average of democracy of the most capable member when the alliance formed differs across conditions on military support and treaty depth. In Figure 4.2, each quadrant corresponds to a particular combination of treaty depth and conditionality. To divide the latent depth measure, I classified deep alliances as treaties with a latent depth score above the median va-

[14] In the GJRM estimator, I use a third equation to model heterogeneity in the error term correlations, which I expect depends on the start year of the alliance. In particular, I suspect that correlations in unobservable factors between treaty depth and unconditional promises of military spending vary over time. Using the start year of the treaty to predict error correlations captures common unobserved shocks from the international context. For example, Kuo (2019) shows how European politics encouraged the proliferation of secret alliances before World War I.

[15] Based on a t-test, the difference between these values is statistically significant.

lue. The leading members of deep and conditional alliances have higher polity scores. Conversely, unconditional alliances with little depth have the lowest average polity scores.

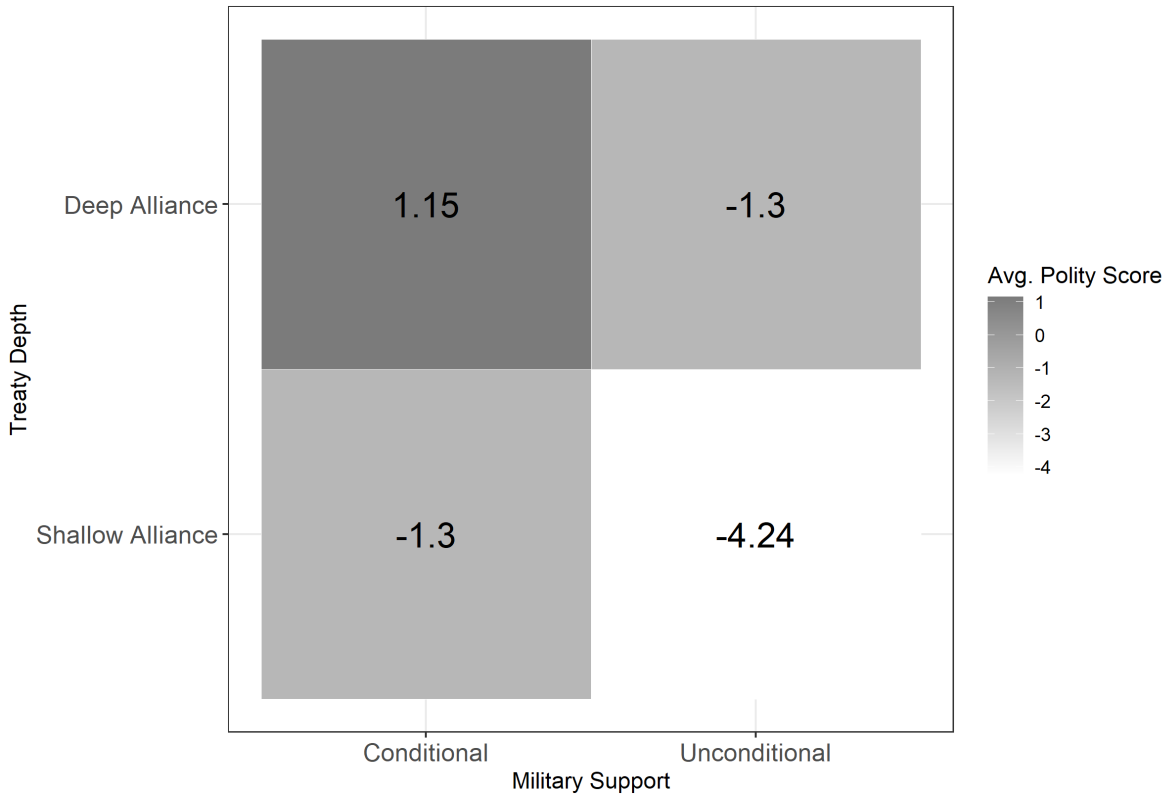

Figure 4.2: Average of the most capable alliance member's polity score when the alliance formed in four groups of alliance from 1816 to 2016. Divisions between alliances based on unconditional military support and treaty depth. Darker quadrants mark a higher average democracy score for that group of alliances, and the text in each box gives the precise value.

These descriptive results do not adjust for potential confounding factors, however. I now describe results from statistical models of depth and unconditional military support, starting with separate models. Table 4.1 shows results from a beta model of treaty depth and a binomial model of unconditional military support with a probit link function. The results are partially consistent with the hypotheses. First, I find a positive association between the democracy of the most capable

alliance member and treaty depth. For unconditional military support, the parameter estimate is negative, but the 95% confidence interval for alliance leader democracy includes zero and positive values.

	Dependent variable:	
	Latent Depth (rescaled)	Unconditional Military Support
	beta	*probit*
	(1)	(2)
Most Capable Member POLITY	0.023	−0.018
	(0.002, 0.043)	(−0.047, 0.011)
Foreign Policy Concessions	−0.057	0.021
	(−0.209, 0.096)	(−0.197, 0.239)
Number of Members	0.016	−0.030
	(−0.011, 0.043)	(−0.077, 0.018)
Wartime Alliance	−0.309	−0.954
	(−0.658, 0.040)	(−1.570, −0.338)
Asymmetric Obligations	0.189	0.035
	(−0.155, 0.532)	(−0.467, 0.537)
Asymmetric Capability	0.347	0.651
	(−0.120, 0.814)	(−0.218, 1.520)
Non-Major Only	0.275	1.146
	(−0.228, 0.779)	(0.266, 2.027)
Average Threat	1.248	1.630
	(0.376, 2.120)	(0.295, 2.965)
Foreign Policy Disagreement	0.197	0.394
	(−0.258, 0.653)	(−0.306, 1.094)
Start Year	0.004	0.015
	(0.0004, 0.007)	(0.010, 0.021)
Constant	−8.929	−31.655
	(−15.679, −2.179)	(−43.220, −20.091)
Observations	277	277
Log Likelihood	54.349	−132.467
Note:		95% Confidence Intervals in Parentheses.

Table 4.1: Independent models of alliance treaty depth and unconditional military support in ATOP offensive and defensive alliances from 1816 to 2007.

As the coefficient estimates alone can mislead, I assess the substantive impact of allied democracy in Figure 4.3. This figure plots the estimated marginal effect of the alliance leader's POLITY score on both outcomes. The left-hand plot of Figure 4.3 shows the association between the average democracy of alliance members and the predicted probability of unconditional military support. This relationship is weaker than expected. These results contradict previous findings that democratic alliance membership increases the likelihood of conditional obligations (Mattes, 2012*b*; Chiba, Johnson and Leeds, 2015).

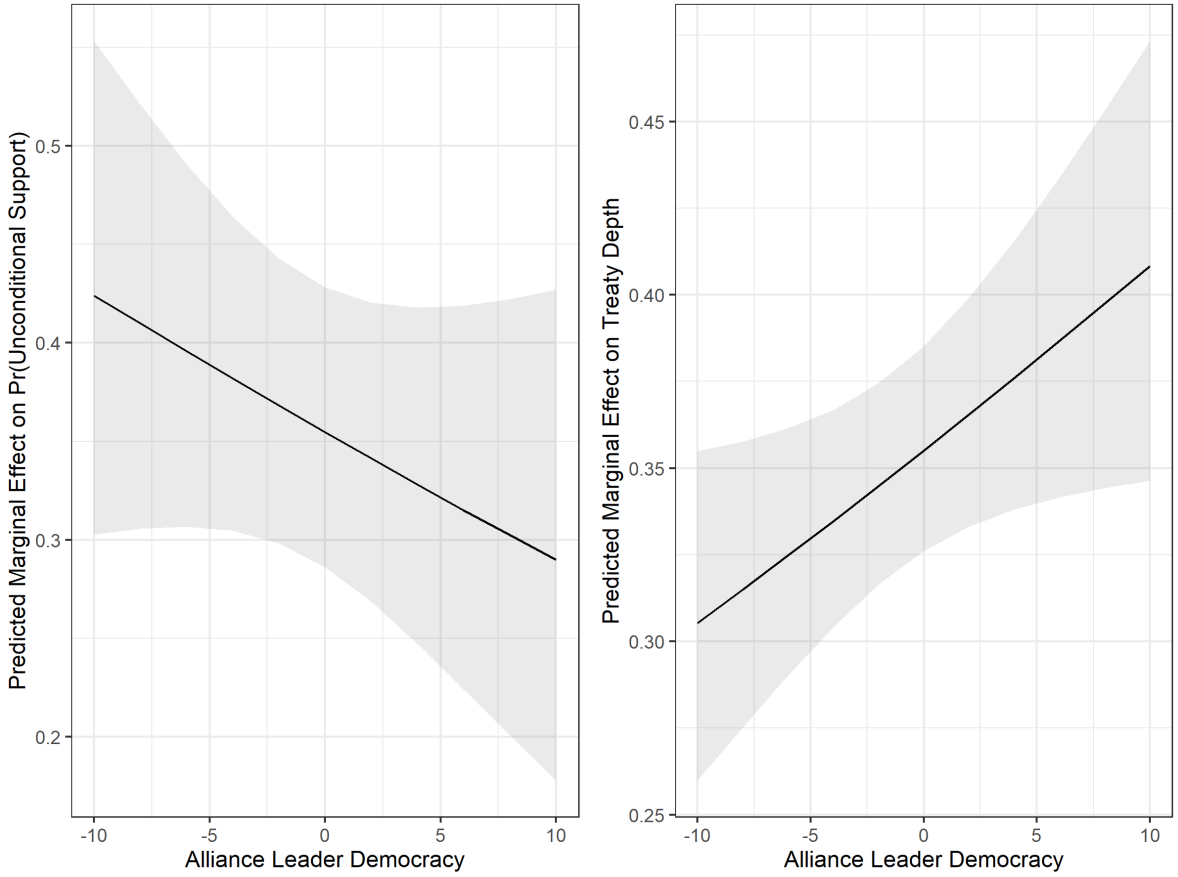

Figure 4.3: Predicted probabilities of unconditional military support and predicted change in treaty depth across the range of average alliance democracy. The line marks predicted values, and the shaded areas encapsulate the standard errors. The rug plot on the x-axis marks observed values of allied democracy. Predictions based on the smoothed terms from a joint generalized regression model.

The right-hand plot in Figure 4.3 shows a positive relationship between average democracy at the time of alliance formation and treaty depth. Substantial democracy in an alliance is associated with greater treaty depth. The predicted value of rescaled treaty depth in an alliance with a fully democratic leader is roughly .1 greater than an alliance with a fully autocratic leader, in expectation. This substantively large relationship matches the treaty depth hypothesis.

Although the results from separate models are informative, they do not account for correlations in the error terms of the depth and unconditional support models, which could affect inferences. I now report the results of the joint analysis of depth and unconditional military support in Table 4.2. This table contains results from both equations of the GJRM model, and marks smoothed terms with the letter s.

As in Table 4.1, I find a positive relationship between the democracy of the leading alliance member and treaty depth, but a weak negative relationship between democratic influence and treaty depth. The bivariate model produces similar inferences. The control variables in this model are also interesting and somewhat different from the estimates in the separate models. Asymmetric capability in an alliance and symmetric alliances between non-major powers are more likely to include unconditional military support than symmetric major power alliances. I also find that the number of members and wartime alliances reduce the probability of unconditional military support. Asymmetric capability and the number of alliance members both increase depth. Last, threat and the year of alliance formation increase unconditional military support and treaty depth.

I make similar inferences about democracy and treaty design with a joint model in Figure 4.3 and the single-equation models in Table 4.1. Inferences about the control variables vary somewhat across the two models, however. I find that the polity score of the most capable alliance member is positively correlated with treaty depth, and has a slight negative association with unconditional military support.

4.3.1 Elections, Political Competition and Executive Constraints

Different institutional characteristics of democracies explain the unconditional military support finding. Having analyzed cumulative democracy scores, I now turn to a model that distinguishes

	Uncond. Mil. Support		Latent Depth	
	Estimate	Std. Error	Estimate	Std. Error
Most Capable POLITY	-0.0173979	0.0158094	0.0262429	0.0097732
Economic Issue Linkage	0.2232389	0.2006227	-0.0060972	0.1458203
FP Concessions	-0.1468542	0.1214414	-0.0339923	0.0845875
Number of Members	-0.0994788	0.0264666	0.0179032	0.0129499
Wartime Alliances	-0.6274650	0.3157751	-0.0744869	0.1787959
Asymmetric Obligations	-0.0181268	0.2622665	0.1686232	0.1632100
Asymmetric Capability	0.9586816	0.3864164	0.3436770	0.2192153
Non-Major Only	1.7040882	0.3975041	0.0828621	0.2310327
FP Disagreement	0.1253382	0.3352438	0.3284441	0.2204690
s(Mean Threat)	7.3253441	43.4564525	1.0000004	16.8780421
s(Start Year)	4.7057429	49.8605879	3.3275472	39.7155651
(Intercept)	-1.0980618	0.4635789	-1.0810691	0.2485382

Table 4.2: Results from joint generalized regression model of treaty depth and unconditional military support. All smoothed terms report the effective degrees of freedom and the chi-squared term. The unconditional military support model is a binomial GLM with a probit link function. The treaty depth model is a beta regression. I model the error correlation between the two processes with a T copula.

between the different parts of the POLITY index. The key independent variables in this model are dummy indicators of competitive elections, open political competition, and executive constraint through parity or subordination to other actors. All three of these factors could contribute to audience costs by exposing leaders to adverse reactions to foreign policy actions. I use a GJRM model to assess this relationship.[16]

The three components of democracy have different consequences for alliance treaty design.[17] First, the presence of competitive elections decreases the probability of unconditional military support and increases treaty depth. This matches the argument— if leaders can be removed from office by competitive elections, they prefer treaty depth to unconditional military support because depth is less salient for voters. Second, the political competition coefficient for treaty depth is negative, albeit with substantial uncertainty. This implies that scrutiny of the executive from open political competition may restrain deep alliance commitments. Political competition has a weak

[16] A T copula has the best model fit and AIC.

[17] These inferences hold the other components of democracy constant.

	Uncond. Mil. Support		Latent Depth	
	Estimate	Std. Error	Estimate	Std. Error
Competitive Elections	-1.3862036	0.5969913	1.0334447	0.3441959
Political Competition	0.1267747	0.4456283	-0.3968825	0.2894709
Executive Constraints	0.7810824	0.2758115	-0.3806161	0.2006471
Economic Issue Linkage	0.0786646	0.1830420	0.1099506	0.1466334
FP Concessions	-0.1284395	0.1256934	-0.1052689	0.0817796
Number of Members	-0.1004983	0.0271633	0.0268154	0.0139085
Wartime Alliances	-0.7495699	0.2448796	0.0599420	0.1700357
Asymmetric Obligations	-0.0230173	0.2251492	0.1766837	0.1620580
Asymmetric Capability	1.0281801	0.5255088	0.4985258	0.2426731
Non-Major Only	1.6291469	0.5180102	0.1175974	0.2551236
FP Disagreement	0.1726706	0.2975962	0.3405041	0.2124881
s(Mean Threat)	7.5870610	44.8092434	1.0000005	25.8169433
s(Start Year)	5.9742148	50.0279652	8.3362695	57.7689902
(Intercept)	-0.8925660	0.6081549	-1.3310577	0.2833844

Table 4.3: Results from joint generalized regression model of treaty depth and unconditional military support, where democracy is divided into competitive elections, open political competition and executive constraints. All smoothed terms report the effective degrees of freedom and the chi-squared term. The unconditional military support model is a binomial GLM with a probit link function. The treaty depth model is a beta regression. I model the error correlation between the two processes with a T copula.

association with the probability of unconditional military support, however.

Last, executive constraints has the opposite effect of electoral competition on both credibility sources, which is interesting and unexpected. First, I find that executive constraints reduce treaty depth. Constraints might reduce treaty depth by exposing the executive to scrutiny from other political elites who have ample information about foreign policy. Treaty depth may then be constrained by political elites that want to avoid foreign entanglements. I also find that presence of domestic institutions with executive parity or subordination increases the probability of unconditional military support. One explanation for the positive relationship between constraints and unconditional support is that democratic leaders want to commit successors with different foreign policy preferences to the alliance Mattes (2012*a*). Leadership turnover in democracies threatens international cooperation, because new leaders may have different supporters and preferences (Lobell, 2004; Narizny, 2007; Leeds, Mattes and Vogel, 2009). By making an unconditional promise of military

support, leaders generate higher audience costs for successor governments to break the alliance.

Taken together, the results from dividing democracy into three core concepts in Table 4.3 help explain the results from the aggregate democracy measure. The large positive effect of competitive elections drives positive association between democracy and treaty depth, as it overwhelms the smaller negative effect of executive constraints. Competitive elections decrease the probability of unconditional military support, but executive constraints make unconditional support an attractive way to precommit future leaders to the alliance. Because these competing effects cancel each other out, there is a weak association between the democracy of the alliance leader and the probability of unconditional military support.

These statistical results give mixed evidence for the argument. Democracies do offer deeper treaties, but democratic institutions have competing effects on the probability of unconditional military support. To further examine my claim that democracies often use treaty depth to reassure, but are less inclined to offer unconditional support, I describe how US preferences shaped the institutional design of NATO.

4.3.2 NATO Treaty Design

I focus this brief case study on NATO treaty design for two reasons. First, information from NATO applies to multiple alliances, as other US treaties have similar designs. Most US alliances have conditional promises of military support and understanding why is important. Second, NATO is also the most important alliance in international politics, so understanding how the treaty formed is worthwhile.

After the end of World War II, the US sought a way to protect Europe from the USSR. Despite acute security concerns, fear of entrapment in unwanted conflicts led to limits on military support. First, as Poast (2019*a*) details, NATO members disagreed over how to define the North Atlantic area, which was a key condition on military support. The US and other states argued about whether France's Algerian colony and Italy should be protected by the alliance. Second, active military support from NATO members depends on domestic political processes.[18] Isolationists in the US

[18]Benson (2012) calls this kind of commitment a "probablistic" obligation.

Senate feared that an alliance would force America to intervene automatically if partners were attacked, bypassing the power of Congress to declare war and engaging the US in unwanted conflicts (Acheson, 1969, pg. 280-1). Therefore Article V of the NATO treaty states that if one member is attacked the others "will assist the Party or Parties so attacked by taking forthwith, individually and in concert with the other Parties, *such action as it deems necessary* (emphasis mine)." Military support was and is not guaranteed. Secretary of State Dean Acheson stated as much in a March 1949 press release defending NATO to the US public, where he said that Article V "does not mean that the United States would automatically be at war if one of the nations covered by the Pact is subject to armed attack" (Acheson, 1949). This claim and the emphases of the press release shows that promises of military support were highly salient to the US public.

Military support from Article V did not assuage European fears that if the Soviets invaded, the United States would not fight. To increase the credibility of NATO, the United States took other measures. A 1951 presentation by Dean Acheson to Dwight Eisenhower argued European allies "fear the inconstancy of United States purpose in Europe. ... These European fears and apprehensions can only be overcome if we move forward with determination and if we make the necessary full and active contribution in terms of both military forces and economic aid" (Acheson, 1951, pg. 3).

The first part of reassurance was the creation of the Atlantic Council, which is an international organization and the main source of depth in the NATO treaty itself. The United States used the Atlantic Council to coordinate collective defense and increase the perceived reliability of the alliance. By investing in the Atlantic Council and related joint military planning, the US addressed European fears of abandonment. For example, US officials thought that the British Foreign Minister viewed US provision of a supreme commander in Europe as "a stimulus to European action" in NATO (Acheson, 1950).

Many Senators also opposed military aid to Europe (Acheson, 1969, pg 285), which limited efforts to add further treaty depth. These legislative constraints on the executive branch reduced the formal depth of NATO relative to what many ambassadors preferred (Acheson, 1969, pg 277).

Bilateral agreements on troop deployments then became another instrument of reassurance. In 1950 the Germans formally requested clarification on whether an attack on US forces in Germany would be treated as an armed attack on the US- which the US said it would (Acheson, 1969, pg. 395). These bilateral arrangements and basing rights are not covered in the NATO treaty, but they added substantial depth.[19]

NATO negotiations reveal the tendency of democracies to use treaty depth to reassure their allies, rather than unconditional military support. Fear of foreign entanglement led the United States to offer conditional military support, but did not inhibit deep military cooperation, which helped reassure European allies. Limits on the promises of military support were an important public justification for the NATO treaty, while the Atlantic Council was less discussed. Still, the power of treaty ratification in the Senate limited formal NATO depth to the Atlantic Council. The Atlantic Council and associated bureaucratic machinery are the formal core of substantial defense cooperation. Altogether though Article V is limited, the US used treaty depth to increase the credibility of NATO.

4.4 Discussion and Conclusion

The findings from the statistical models and case study of NATO generate mixed evidence for the hypotheses. I find regular evidence that democracies often form deep alliances, which may be driven by selecting leaders through open elections. There is inconsistent support for claims that allied democracy decreases the probability of unconditional military support, however, because competitive elections and executive constraints have contradictory effects. Elections encourage conditional obligations, but executive constraints may encourage locking successors into the alliance with unconditional support.

The results are mostly consistent with my overarching claim states form deep alliances to increase the credibility of their alliance commitments while managing the risk of entrapment and audience costs. Because democracies are concerned with audience costs and treaty depth has little salience with voters, they often use treaty depth to increase the credibility of their alliances.

[19]This is a potential limitation of the statistical models.

Electoral politics are largely responsible for this relationship, and they also push democracies away from unconditional military support.

My argument and evidence have two limitations. For one, I only examine variation in formal treaty design. This omits the implementation of alliance promises, which may be deeper or shallower than the treaty language alone implies. As the NATO case study shows, formal treaty depth reflects practical depth, but it may miss some differences between alliances. Changes in realized alliance depth are a useful subject for future inquiry, but will require new data collection. Also, the small sample size of observed alliances adds uncertainty to inferences.

Despite its limitations, this paper has four main implications for scholarship. First, one reason treaty depth matters is that it affects military spending by alliance participants. The findings in this paper imply that states do not use treaty depth to manipulate allied military spending, but rather to increase the credibility of the alliance. Therefore, treaty depth is non-randomly selected based on observable alliance characteristics like domestic institutions, but selection on unobservable preferences over allied military spending is less likely.

Second, studies of how alliance participation affects international politics must account for alliance design and membership. Alliance member characteristics and treaty design are correlated, and both affect the consequences of treaty participation. Estimating the impact of member characteristics alone, or treaty design alone, risks omitted variable bias.

The third implication is that democracies do not make limited alliance commitments. Even if democracies impose conditions on military support, deep alliances add substantial foreign entanglement. The most limited alliances possible would have conditional obligations and no depth. Last, some of the lessons from this work add to the extensive literature on the design on international institutions (Downs and Rocke, 1995; Martin and Simmons, 1998; Koremenos, Lipson and Snidal, 2001; Koremenos, 2005; Thompson, 2010). Just as democracies use depth to support allies while managing electoral risks, democracies may undertake international commitments in ways that limit public scrutiny.

The findings also raise at least two questions for future research. First, they address debates

about whether democracies make more credible commitments than other states. The effect of democracy on credibility can be divided into conditions on military support, treaty depth, and the direct effect of institutions and domestic politics. These may have competing or conditional effects, which could explain mixed findings about the credibility of democratic commitments (Schultz, 1999; Leeds, 1999; Thyne, 2012; Downes and Sechser, 2012). Future research should combine the components of democracy and democratic alliances to asses the net effect of democracy on credible commitment in international relations.

Scholars should also consider how alliance treaty design varies among different types of autocracies. As I noted in the argument, some autocratic states have high audience costs for backing down in military interventions. Differences in the salience of audience costs, which actors impose audience costs on leaders and what information those actors have about foreign policy (Weeks, 2008) may help explain alliance treaty design. For example, personalist leaders with few public or elite constraints on their foreign policy may be able to form alliances with depth and unconditional military support.

In conclusion, states use deep alliances to reassure their partners while limiting the risk of entrapment. Domestic political institutions shape how states build the credibility of their alliances. Thanks to high audience costs of military intervention, democracies are especially likely to use treaty depth to increase the credibility of their alliances.

5. SUMMARY AND CONCLUSIONS

5.1 What Did I Do?

This dissertation addressed the question of alliance participation and military spending. First, I reassessed an alternative to my bargaining-focused argument— the public goods model of alliances. After noting that existing empirical tests of the public goods logic suffer from model specification and generalizability problems, I offered an improved empirical test. My examination of the association between economic weight in alliances and percentage changes in military spending found little evidence that small states decrease military spending and large states increase military spending due to alliance participation. In the second chapter, I showed that deep alliance treaties often decrease non-major power military spending, but participation in shallow alliances is more likely to increase non-major power military spending. Last, I explored the sources of alliance treaty depth. I argued that states use treaty depth to assuage allied concerns about treaty reliability in a way that minimizes the domestic audience costs of international commitment. Because democracies have higher audience costs from military support, they use treaty depth to increase alliance credibility. I then showed that democracies are more likely to form deep alliances using a statistical analysis and illustrative case study of NATO.

5.2 What Did We Learn?

We learned four things from this dissertation project. First, we learned that alliance participation can increase or decrease military spending. Second, we learned that a crucial prediction of free-riding in alliances based on economic size has limited empirical support. Third, we learned that the impact of alliance participation on military spending depends on the relative bargaining leverage of alliance members difference alliance members state capability and alliance treaty design. Therefore, claims that alliance participation only increases or only decreases military spending are incomplete, because each applies in a specific set of circumstances. Last, we learned that deep treaties follow from attempts to establish a credible commitment of military support between alliance

partners while managing the risk of entrapment.

All told, my findings suggest that bargaining between alliance members and the credibility of commitment shape the connection between alliance participation and military spending. States' leverage over allied military spending depends on how they design alliances in response to abandonment and entrapment concerns. Put differently, efforts to ensure credible commitment through alliance treaty design affect alliance members' military spending.

5.3 Why Does it Matter?

My argument and findings contribute to scholarship on alliance politics and international relations. I first address two implications for alliance politics scholarship. After that, I turn to two implications for research on international institutions and international relations more generally.

In alliance politics, the conventional wisdom about non-major powers, alliances, and military spending should be revised. While conventional wisdom expects regular non-major power free-riding, I find that alliance participation often increases or has little effect on non-major power military spending. Some alliances do reduce non-major power military spending, but these alliances are unusually deep and capable.

The findings about when non-major powers can reduce military spending in alliances should also inform debates about US alliances. If the United States wants allies to spend more on defense, it may need to change the nature of its alliance commitments. By increasing the credibility of US commitments, deep US alliances may undermine attempts to encourage higher allied defense spending. The United States formed deep alliances to reassure allied states, but reassurance may be incompatible with demanding more investment in military capability.

Therefore, reconciling competing views of alliances in US grand strategy will be difficult. If the United States wants allies to spend more on defense, less emphasis on credible commitments may be necessary. If allies have low defense spending, deep alliances like NATO rarely encourage increases. This means that deep engagement to reassure partners may lead allies to maintain low defense spending, which is a common criticism of US alliances in the restraint school.

In addition to alliance politics, this dissertation makes two contributions to international relati-

ons scholarship. First, the results provide new evidence about how how states use different policy tools as complements and substitutes. Arms and alliances are a classic example of substitutes, but I find that they are complements for non-major powers in shallow alliances and substitutes in deep alliances. Therefore, even if policies provide the same good, substitution is unusual. It may be hard to find empirical evidence of substitution in foreign policy (Starr, 2000) because substitution is rare and highly conditional.

Second, the argument and findings in Chapter 3 reinforce the importance of accounting for differences in institutional design. Studies of trade e.g. (Allee and Scalera, 2012; Tomz, Goldstein and Rivers, 2007), bilateral investment treaties (Blake, 2013; Simmons, 2014) and the rational design of institutions literature (Koremenos, Lipson and Snidal, 2001) all acknowledge this point. Translating theories of institutional heterogeneity into research design is challenging, however. Many studies of international institutions rely on state-level aggregate measures. When scholars are interested in inferring how institutional characteristics modify the consequences of participation, they could employ variants of the multilevel model in Chapter 3. If scholars are only interested in estimating the specific effects of individual institutions, and states are not neatly nested within institutions, the model in Chapter 2 could facilitate inferences.

5.4 What's Next?

To extend the dissertation research, I first plan to write a book on alliance participation and military spending. This dissertation offers a preliminary answer to the question of alliance participation and military spending. Chapter 2 deals with non-major powers and treaty depth, but major powers are important actors in international politics. Moreover, other alliance characteristics alter the connection between alliances and military expenditures.

In the book, I will argue alliance participation can increase, decrease, or have no effect on military expenditures, depending on state capability and the characteristics of the alliance. Capability shapes whether a state uses alliances and military spending primarily for security or influence. Major powers are the most capable states, and they use alliances for influence— reshaping international relations in beneficial ways. Less capable non-major powers use alliances for immediate

territorial security. As a result, major and non-major power military spending responds differently to alliance characteristics including treaty design and membership.

The book project will consider multiple alliance characteristics. Besides treaty depth, I will focus on foreign policy concessions, issue linkages in alliance treaties, alliance size, and foreign policy similarity. Among non-major powers, my argument emphasizes the relationship between reassurance and free-riding. The major power argument will include credibility concerns, but also consider how alliance participation affects foreign obligations. To give a preliminary example, I expect that adding economic issue linkages to an alliance increases the impact of alliance participation on major power military spending, but does not modify the association between alliance participation and non-major power defense spending. This occurs because issue linkages can overcome bargaining challenges and encourage major powers to increase their foreign obligations. Non-major powers in treaties with issue linkages benefit from greater credibility, but the linked issues are a source of allied leverage over free-riding, so linkages have no effect on military spending.

Undertaking a book project will also allow me to include a wider range of empirical evidence. This will include a multilevel model for major powers, and a model that makes inferences about major and non-major powers at the same time. A varying slopes multilevel model or a multilevel model with state-specific alliance parameters could also distinguish between major and non-major powers, as well as major powers and superpowers. Placing major powers alongside non-major powers will provide a comprehensive theoretical and empirical framework for understanding alliances and military spending.

To complement the book project, I plan to develop a research agenda on the domestic politics of international alliances. There is already some research on this subject e.g. (Narizny, 2003; Lobell, 2004; Leeds, Mattes and Vogel, 2009), but there are many open questions. Alliance participation and military spending are both closely connected to domestic politics. As Chapter 4 showed, domestic political institutions shape alliance treaty design, which has consequences for military spending.

Alliance participation may also change domestic politics, however. Changes in defense spending from alliance participation may affect the domestic economy and the "guns-butter tradeoff" (Whitten and Williams, 2011; Poast, 2019*b*). By changing patterns of military spending and trade (Gowa and Mansfield, 2004), alliances could create their own domestic constituencies. Understanding the sources of domestic support for international alliances is a potentially interesting subject for future research.

Beyond the material consequences of changes in military spending, changes in military spending by allied states may also affect public opinion about international security cooperation. Low allied defense spending could reduce public support for alliance participation. But public support may depend on how allied military spending decisions are presented. There are two ways to frame low allied defense spending, which reflect existing theories. Low defense spending could be called "free-riding," or portrayed as the outcome of bargaining and exchange between alliance members.

A collective action frame defines low allied spending in terms of the common good and individual incentives to defect. Insufficient allied defense spending violates the purpose and function of the alliance. As a result, a collective action frame activates a fear of exploitation and conditional cooperation norms, which could reduce support for alliance participation. An alternative frame treats low allied spending as part of a mutually beneficial exchange. If a patron state provides protection in exchange for foreign policy concessions, their proteges can lower defense spending (Morrow, 1991; Lake, 2009). Here, low allied defense spending reflects the purpose and function of the alliance. Exchanging protection for autonomy benefits alliance members in different ways and is less likely to raise a sense of exploitation. Therefore, bargaining framing of alliances are less likely to reduce support for alliance participation.

To examine how framing allied military spending affects domestic support, I plan to field a survey experiment on the US public. The experiment would treat audiences with free-riding, exchange and neutral frames of allied military spending and examine how support for remaining in the alliance, sanctioning allies, and military intervention to protect allies changes. The neutral frame would be the base category for the exchange and collective action frames. I already secured

some funding for this experiment.

A related study could examine how the US media portrays alliances. Chapter 3 assumes that the public has little information on alliances outside of crises where treaty participation might involve a country in war. In general, alliances are not a salient political issue, though growing debate about America's role in the world may be changing that. Observational data on the content of US media, such as TV transcripts from the Chapel Hill American Media Project, might provide insight into what the public might know about NATO and other alliances.

REFERENCES

Acheson, Dean. 1949. "State Department Press Release, March 18, 1949.".

Acheson, Dean. 1950. "Telegram, Dean Acheson to Harry S. Truman, September 14, 1950.".

Acheson, Dean. 1951. "Dean Acheson to Harry S. Truman, with attachment, January 5, 1951." Memo.

Acheson, Dean. 1969. *Present at the Creation*. New York: W.W. Norton.

Allee, Todd L. and Jamie E. Scalera. 2012. "The Divergent Effects of Joining International Organizations: Trade Gains and the Rigors of WTO Accession." *International Organization* 66(2):243–276.

Allen, Michael A and Matthew DiGiuseppe. 2013. "Tightening the Belt: Sovereign Debt and Alliance Formation." *International Studies Quarterly* 57(4):647–659.

Altfield, Michael F. 1984. "The Decision to Ally: A Theory and Test." *Western Political Quarterly* 37(4):523–44.

Barnett, Michael N and Jack S Levy. 1991. "Domestic Sources of Alliances and Alignments: The Case of Egypt, 1962-73." *International Organization* 45(3):369–395.

Barrett, Scott. 2010. *Why Cooperate?: The Incentive to Supply Global Public Goods*. Oxford University Press.

Beckley, Michael. 2015. "The Myth of Entangling Alliances: Reassessing the Security Risks of U.S. Defense Pacts." *International Security* 39(4):7–48.

Benson, Brett V. 2012. *Constructing International Security: Alliances, Deterrence, and Moral Hazard*. Cambridge University Press.

Benson, Brett V and Joshua D Clinton. 2016. "Assessing the Variation of Formal Military Alliances." *Journal of Conflict Resolution* 60(5):866–898.

Berkemeier, Molly and Matthew Fuhrmann. 2018. "Reassessing the fulfillment of alliance commitments in war." *Research & Politics* April-June:1–5.

Betancourt, Michael and Mark Girolami. 2015. Hamiltonian Monte Carlo for hierarchical models.

In *Current Trends in Bayesian Methodology with Applications*, ed. Satyanshu K. Upadhyay, Umesh Singh, Dipak K. Dey and Appaia Loganathan. Chapman and Hall/CRC Press pp. 79–102.

Blake, Daniel J. 2013. "Thinking Ahead: Government Time Horizons and the Legalization of International Investment Agreements." *International Organization* 67(3):797–827.

Bolt, J, R Inklaar, H de Jong and JL van Zanden. 2018. "Maddison Project Database, Version 2018." *Rebasing 'Maddison': new income comparisons and the shape of long-run economic development* .

Boyer, Mark A. 1993. *International Cooperation and Public Goods: Opportunities for the Western Alliance*. Baltimore: Johns Hopkins University Press.

Brands, Hal and Peter D. Feaver. 2017. "What are America's Alliances Good for?" *Parameters* 47(2):15–30.

Braumoeller, Bear F, Giampiero Marra, Rosalba Radice and Aisha E Bradshaw. 2018. "Flexible causal inference for political science." *Political Analysis* 26(1):54–71.

Brooks, Stephen G, G John Ikenberry and William C Wohlforth. 2013. "Don't come home, America: the case against retrenchment." *International Security* 37(3):7–51.

Carpenter, Bob, Andrew Gelman, Matt Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Michael A Brubaker, Jiqiang Guo, Peter Li and Allen Riddell. 2016. "Stan: A probabilistic programming language." *Journal of Statistical Software* 20:1–37.

Chen, Baizhu, Yi Feng and Cyrus Masroori. 1996. "Collective action in the Middle East? A study of free-ride in defense spending." *Journal of Peace Research* 33(3):323–339.

Chiba, Daina, Jesse C Johnson and Brett Ashley Leeds. 2015. "Careful Commitments: Democratic States and Alliance Design." *The Journal of Politics* 77(4):968–982.

Clinton, Joshua, Simon Jackman and Douglas Rivers. 2004. "The Statistical Analysis of Roll Call Data." *American Political Science Review* 98(2):355–370.

Conybeare, John AC. 1994. "Arms versus Alliances: The Capital Structure of Military Enterprise." *Journal of Conflict Resolution* 38(2):215–235.

Conybeare, John AC and Todd Sandler. 1990. "The Triple Entente and the Triple Alliance 1880–1914: A Collective Goods Approach." *American Political Science Review* 84(4):1197–1206.

Crescenzi, Mark JC, Jacob D Kathman, Katja B Kleinberg and Reed M Wood. 2012. "Reliability, Reputation, and Alliance Formation." *International Studies Quarterly* 56(2):259–274.

Davis, Christina L. 2004. "International Institutions and Issue Linkage: Building Support for Agricultural Trade Liberalization." *American Political Science Review* 98(1):153–169.

Diehl, Paul F. 1994. "Substitutes or Complements?: The Effects of Alliances on Military Spending in Major Power Rivalries." *International Interactions* 19(3):159–176.

DiGiuseppe, Matthew and Paul Poast. 2016. "Arms versus Democratic Allies." *British Journal of Political Science* pp. 1–23.

Downes, Alexander B and Todd S Sechser. 2012. "The Illusion of Democratic Credibility." *International Organization* 66(3):457–489.

Downs, George W. and David M. Rocke. 1995. *Optimal Imperfection?: Domestic Uncertainty and Institutions in International Relations*. Princeton, New Jersey: Princeton University Press.

Fariss, Christopher J. 2014. "Respect for Human Rights has Improved Over Time: Modeling the Changing Standard of Accountability." *American Political Science Review* 108(2):297–318.

Fearon, James D. 1997. "Signaling Foreign Policy Interests: Tying Hands versus Sinking Costs." *Journal of Conflict Resolution* 41(1):68–90.

Fearon, James D. 1998*a*. "Bargaining, enforcement, and international cooperation." *International Organization* 52(2):269–305.

Fearon, James D. 1998*b*. "Domestic Politics, Foreign Policy, and Theories of International Relations." *Annual Review of Political Science* 1:289–313.

Fearon, James D. 2018. "Cooperation, Conflict, and the Costs of Anarchy." *International Organization* pp. 1–37.

Fjelstul, Joshua C and Dan Reiter. 2019. "Explaining incompleteness and conditionality in alliance agreements." *International Interactions* 45(6):976–1002.

Fordham, Benjamin. 1998. *Building the Cold War Consensus: The Political Economy of US*

National Security Policy, 1949-51. Ann Arbor: University of Michigan Press.

Fordham, Benjamin O. 2010. "Trade and asymmetric alliances." *Journal of Peace Research* 47(6):685–696.

Fordham, Benjamin and Paul Poast. 2014. "All Alliances Are Multilateral Rethinking Alliance Formation." *Journal of Conflict Resolution* 58(27):1–26.

Fuhrmann, Matthew. 2020. "When Do Leaders Free-Ride? Business Experience and Contributions to Collective Defense." *American Journal of Political Science* forthcoming.

Fuhrmann, Matthew and Todd S Sechser. 2014. "Signaling Alliance Commitments: Hand-Tying and Sunk Costs in Extended Nuclear Deterrence." *American Journal of Political Science* 58(4):919–935.

Garfinkel, Michelle R. 2004. "Stable alliance formation in distributional conflict." *European Journal of Political Economy* 20(4):829–852.

Gartzke, Erik and Kristian Skrede Gleditsch. 2004. "Why democracies may actually be less reliable allies." *American Journal of Political Science* 48(4):775–795.

Gates, William R and Katsuaki L Terasawa. 1992. "Commitment, Threat Perceptions, and Expenditures in a Defense Alliance." *International Studies Quarterly* 36(1):101–118.

Gaubatz, Kurt Taylor. 1996. "Democratic states and commitment in international relations." *International Organization* 50(1):109–139.

Gavin, Francis J. 2015. "Strategies of inhibition: US grand strategy, the nuclear revolution, and nonproliferation." *International Security* 40(1):9–46.

Gelman, Andrew, Daniel Simpson and Michael Betancourt. 2017. "The prior can generally only be understood in the context of the likelihood." *arXiv preprint arXiv:1708.07487* .

Gelman, Andrew and Jennifer Hill. 2007. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Vol. 1 Cambridge University Press New York, NY, USA.

George, Justin and Todd Sandler. 2017. "Demand for military spending in NATO, 1968–2015: A spatial panel approach." *European Journal of Political Economy* 53:222–236.

Gibler, Douglas M. 2008. "The Costs of Reneging: Reputation and Alliance Formation." *Journal*

of Conflict Resolution 52(3):426–454.

Gibler, Douglas M and Scott Wolford. 2006. "Alliances, Then Democracy: An Examination of the Relationship Between Regime Type and Alliance Formation." *Journal of Conflict Resolution* 50(1):129–153.

Gibler, Douglas M, Steven V Miller and Erin K Little. 2016. "An Analysis of the Militarized Interstate Dispute (MID) Dataset, 1816–2001." *International Studies Quarterly* 60(4):719–730.

Goldsmith, Benjamin E. 2003. "Bearing the Defense Burden, 1886-1989: Why Spend More?" *Journal of Conflict Resolution* 47(5):551–573.

Goldstein, Avery. 1995. "Discounting the free ride: alliances and security in the postwar world." *International Organization* 49(1):39–71.

Gowa, Joanne. 1995. *Allies, Adversaries, and International Trade*. Princeton University Press.

Gowa, Joanne and Edward D. Mansfield. 2004. "Alliances, Imperfect Markets, and Major-Power Trade." *International Organization* 58(4):775–805.

Granger, Clive WJ and Paul Newbold. 1974. "Spurious Regressions in Econometrics." *Journal of Econometrics* 2(2):111–120.

Häge, Frank M. 2011. "Choice or circumstance? Adjusting measures of foreign policy similarity for chance agreement." *Political Analysis* 19(3):287–305.

Hilton, Brian and Anh Vu. 1991. "The McGuire model and the economics of the NATO alliance." *Defense and Peace Economics* 2(2):105–121.

Horowitz, Michael C, Paul Poast and Allan C Stam. 2017. "Domestic Signaling of Commitment Credibility: Military Recruitment and Alliance Formation." *Journal of Conflict Resolution* 61(8):1682–1710.

Johnson, Jesse C. 2015. "The cost of security: Foreign policy concessions and military alliances." *Journal of Peace Research* 52(5):665–679.

Johnson, Jesse C., Brett Ashley Leeds and Ahra Wu. 2015. "Capability, Credibility, and Extended General Deterrence." *International Interactions* 41(2):309–336.

Juárez, Miguel A and Mark FJ Steel. 2010. "Model-Based Clustering of Non-Gaussian Panel Data

Based on Skew-t Distributions." *Journal of Business & Economic Statistics* 28(1):52–66.

Kim, Tongfi. 2011. "Why Alliances Entangle But Seldom Entrap States." *Security Studies* 20(3):350–377.

Kim, Tongfi. 2016. *The Supply Side of Security : A Market Theory of Military Alliances*. Stanford University Press.

Kim, Wukki and Todd Sandler. 2019. "NATO at 70: Pledges, Free Riding, and Benefit-Burden Concordance." *Defence and Peace Economics* pp. 1–14.

Kimball, Anessa L. 2010. "Political Survival, Policy Distribution, and Alliance Formation." *Journal of Peace Research* 47(4):407–419.

Kono, Daniel Y. 2006. "Optimal Obfuscation: Democracy and Trade Policy Transparency." *American Political Science Review* 100(03):369–384.

Koremenos, Barbara. 2005. "Contracting around International Uncertainty." *American Political Science Review* 99(04):549–565.

Koremenos, Barbara, Charles Lipson and Duncan Snidal. 2001. "The Rational Design of International Institutions." *International Organization* 55(04):761–799.

Kuo, Raymond. 2019. "Secrecy among Friends: Covert Military Alliances and Portfolio Consistency." *Journal of Conflict Resolution* pp. 1–27.

Lai, Brian and Dan Reiter. 2000. "Democracy, Political Similarity, and International Alliances, 1816—1992." *Journal of Conflict Resolution* 44(2):203–27.

Lake, David A. 1996. "Anarchy, Hierarchy and the Variety of International Relations." *International Organization* 50(1):1–33.

Lake, David A. 2009. *Hierarchy in International Relations*. Ithaca: Cornell University Press.

Lanoszka, Alexander. 2015. "Do Allies Really Free Ride?" *Survival* 57(3):133–152.

Leeds, Brett Ashley. 1999. "Democratic Political Institutions, Credible Commitments, and International Cooperation." *American Journal of Political Science* 43(4):979–1002.

Leeds, Brett Ashley. 2003*a*. "Alliance Reliability in Times of War: Explaining State Decisions to Violate Treaties." *International Organization* 57(4):801–827.

Leeds, Brett Ashley. 2003*b*. "Do Alliances Deter Aggression? The Influence of Military Alliances on the Initiation of Militarized Interstate Disputes." *American Journal of Political Science* 47(3):427–439.

Leeds, Brett Ashley, Andrew G Long and Sara McLaughlin Mitchell. 2000. "Reevaluating Alliance Reliability: Specific Threats, Specific Promises." *Journal of Conflict Resolution* 44(5):686–699.

Leeds, Brett Ashley and Burcu Savun. 2007. "Terminating Alliances: Why Do States Abrogate Agreements?" *The Journal of Politics* 69(4):1118–1132.

Leeds, Brett Ashley, Michaela Mattes and Jeremy S Vogel. 2009. "Interests, institutions, and the reliability of international commitments." *American Journal of Political Science* 53(2):461–476.

Leeds, Brett Ashley and Sezi Anac. 2005. "Alliance Institutionalization and Alliance Performance." *International Interactions* 31(3):183–202.

Leeds, Brett, Jeffrey Ritter, Sara Mitchell and Andrew Long. 2002. "Alliance Treaty Obligations and Provisions, 1815-1944." *International Interactions* 28(3):237–260.

Levendusky, Matthew S and Michael C Horowitz. 2012. "When Backing Down is the Right Decision: Partisanship, New Information, and Audience Costs." *The Journal of Politics* 74(2):323–338.

Li, Quan. 2003. "The effect of security alliances on exchange-rate regime choices." *International Interactions* 29(2):159–193.

Lobell, Steven E. 2004. "Politics and national security: The battles for Britain." *Conflict Management and Peace Science* 21(4):269–286.

Long, Andrew G. 2003. "Defense Pacts and International Trade." *Journal of Peace Research* 40(5):537–552.

Long, Andrew G and Brett Ashley Leeds. 2006. "Trading for Security: Military Alliances and Economic Agreements." *Journal of Peace Research* 43(4):433–451.

Martin, Lisa L and Beth A Simmons. 1998. "Theories and Empirical Studies of International Institutions." *International Organization* 52(4):729–757.

Mattes, Michaela. 2012*a*. "Democratic reliability, precommitment of successor governments, and

the choice of alliance commitment." *International Organization* 66(1):153–172.

Mattes, Michaela. 2012*b*. "Reputation, Symmetry, and Alliance Design." *International Organization* 66(4):679–707.

Mattes, Michaela and Mariana Rodríguez. 2014. "Autocracies and international cooperation." *International Studies Quarterly* 58(3):527–538.

McElreath, Richard. 2016. *Statistical Rethinking: A Bayesian course with examples in R and Stan*. Vol. 122 CRC Press.

McManus, Roseanne W and Keren Yarhi-Milo. 2017. "The Logic of "Offstage" Signaling: Domestic Politics, Regime Type, and Major Power-Protégé Relations." *International Organization* 71(4):701–733.

Mearsheimer, John J. 1994. "The False Promise of International Institutions." *International Security* 19(3):5–49.

Morgan, T Clifton and Glenn Palmer. 2006. *A Theory of Foreign Policy*. Princeton: Princeton University Press.

Morgenthau, Hans. 1948. *Politics Among Nations: The Struggle for Power and Peace*. New York: Alfred A. Knopf.

Morrow, James D. 1991. "Alliances and Asymmetry: An Alternative to the Capability Aggregation Model of Alliances." *American Journal of Political Science* 35(4):904–933.

Morrow, James D. 1993. "Arms versus allies: trade-offs in the search for security." *International Organization* 47(2):207–233.

Morrow, James D. 1994. "Alliances, Credibility, and Peacetime Costs." *Journal of Conflict Resolution* 38(2):270–297.

Morrow, James D. 2000. "Alliances: Why Write Them Down?" *Annual Review of Political Science* 3:63–83.

Most, Benjamin A and Harvey Starr. 1989. *Inquiry, Logic and International Politics*. Columbia: University of South Carolina Press.

Most, Benjamin A. and Randolph M. Siverson. 1987. Substituting Arms and Alliances 1870–1914:

an Exploration in Comparative Foreign Policy. In *New Directions in the Study of Foreign Policy*, ed. Charles F Hermann, Charles W. Kegley Jr and James N. Rosenau. Boston: Allen & Unwin pp. 131–160.

Murray, Jared S, David B Dunson, Lawrence Carin and Joseph E Lucas. 2013. "Bayesian Gaussian Copula Factor Models for Mixed Data." *Journal of the American Statistical Association* 108(502):656–665.

Narizny, Kevin. 2003. "Both Guns and Butter, or Neither: Class Interests in the Political Economy of Rearmament." *American Political Science Review* 97(2):203–220.

Narizny, Kevin. 2007. *The Political Economy of Grand Strategy*. Ithaca: Cornell University Press.

Niou, Emerson MS and Sean M Zeigler. 2019. "External Threat, Internal Rivalry, and Alliance Formation." *The Journal of Politics* 81(2):571–584.

Nordhaus, William, John R Oneal and Bruce Russett. 2012. "The Effects of the International Security Environment on National Military Expenditures: A Multicountry Study." *International Organization* 66(3):491–513.

Norrlof, Carla. 2010. *America's Global Advantage: US Hegemony and International Cooperation*. Cambridge University Press.

Olson, Mancur and Richard Zeckhauser. 1966. "An Economic Theory of Alliances." *The Review of Economics and Statistics* 48(3):266–279.

Oneal, John R. 1990. "The theory of collective action and burden sharing in NATO." *International Organization* 44(3):379–402.

Oneal, John R and Hugh Carter Whatley. 1996. "The effect of alliance membership on national defense burdens, 1953–88: A test of Mancur Olson's theory of collective action." *International Interactions* 22(2):105–122.

Palmer, Glenn. 1990. "Alliance Politics and Issue Areas: Determinants of Defense Spending." *American Journal of Political Science* 34(1):190–211.

Pamp, Oliver, Florian Dendorfer and Paul W. Thurner. 2018. "Arm your friends and save on defense? The impact of arms exports on military expenditures." *Public Choice* 177(1):165–87.

Plümper, Thomas and Eric Neumayer. 2015. "Free-riding in alliances: Testing an old theory with a new method." *Conflict Management and Peace Science* 32(3):247–268.

Poast, P. 2012. "Does Issue Linkage Work? Evidence from European Alliance Negotiations, 1860 to 1945." *International Organization* 66(1):277–310.

Poast, Paul. 2013. "Can Issue Linkage Improve Treaty Credibility? Buffer State Alliances as a "Hard Case"." *Journal of Conflict Resolution* 57(5):739–764.

Poast, Paul. 2019*a*. *Arguing about Alliances: The Art of Agreement in Military-Pact Negotiations*. Cornell University Press.

Poast, Paul. 2019*b*. "Beyond the "Sinew of War": The Political Economy of Security as a Subfield." *Annual Review of Political Science* 22:223–239.

Posen, Barry R. 2014. *Restraint: A New Foundation for U.S. Grand Strategy*. Cornell University Press.

Preble, Christopher A. 2009. *The Power Problem: How American Military Dominance Makes Us Less Safe, Less Prosperous, and Less Free*. Ithaca: Cornell University Press.

Pryor, Frederic L. 1968. *Public Expenditures in Communist and Capitalist Nations*. RD Irwin.

Quinn, Kevin M. 2004. "Bayesian Factor Analysis for Mixed Ordinal and Continuous Responses." *Political Analysis* 12(4):338–353.

Quiroz Flores, Alejandro. 2011. "Alliances as Contiguity in Spatial Models of Military Expenditures." *Conflict Management and Peace Science* 28(4):402–418.

Rainey, Carlisle and Daniel K. Baissa. 2020. "When BLUE Is Not Best: Non-Normal Errors and the Linear Model." *Political Science Research & Methods* 8(1):136–148.

Reisinger, William M. 1983. "East European military expenditures in the 1970s: collective good or bargaining offer?" *International Organization* 37(1):143–155.

Reiter, Dan, Allan C. Stam and Michael C. Horowitz. 2016. "A Revised Look at Interstate Wars, 1816–2007." *Journal of Conflict Resolution* 60(5):956–976.

Russett, Bruce. 1970. *What Price Vigilance?: The Burdens of National Defense*. New Haven: Yale University Press.

Sandler, Todd. 1993. "The economic theory of alliances: a survey." *Journal of Conflict Resolution* 37(3):446–483.

Sandler, Todd and John F Forbes. 1980. "Burden sharing, strategy, and the design of NATO." *Economic Inquiry* 18(3):425–444.

Sandler, Todd and Keith Hartley. 2001. "Economics of alliances: The lessons for collective action." *Journal of Economic Literature* 39(3):869–896.

Sarkees, Meredith Reid and Frank Whelon Wayman. 2010. *Resort to War: A Data Guide to Inter-state, Extra-state, Intra-state, and Non-state Wars, 1816-2007*. Washington, DC: CQ Press.

Schuessler, John M and Joshua R Shifrinson. 2019. "The Shadow of Exit from NATO." *Strategic Studies Quarterly* 13(3):38–51.

Schultz, Kenneth A. 1999. "Do Democratic Institutions Constrain or Inform? Contrasting Two Institutional Perspectives on Democracy and War." *International Organization* 53(2):233–266.

Seawright, Jason. 2016. *Multi-Method Social Science: Combining Qualitative and Quantitative Tools*. Cambridge University Press.

Seawright, Jason and John Gerring. 2008. "Case selection techniques in case study research: A menu of qualitative and quantitative options." *Political Research Quarterly* 61(2):294–308.

Senese, Paul D and John A Vasquez. 2008. *The Steps to War: An Empirical Study*. Princeton University Press.

Simmons, Beth A. 2014. "Bargaining over BITS, Arbitrating Awards: The Regime for Protection and Promotion of International Investment." *World Politics* 66(1):12–46.

Simon, Jeffrey. 1985. *Warsaw Pact Forces: Problems of Command and Control*. Boulder, CO: Westview.

Singer, J David. 1988. "Reconstructing the correlates of war dataset on material capabilities of states, 1816–1985." *International Interactions* 14(2):115–132.

Siroky, David S. 2012. "Each Man For Himself? Rival Theories of Alliance Economics in The Early State System." *Defence and Peace Economics* 23(4):321–330.

Smith, Alastair. 1995. "Alliance Formation and War." *International Studies Quarterly* 39(4):405–

425.

Snyder, Glenn H. 1984. "The Security Dilemma in Alliance Politics." *World Politics* 36(04):461–495.

Snyder, Glenn H. 1990. "Alliance theory: A neorealist first cut." *Journal of International Affairs* 44(1):103–123.

Snyder, Glenn H. 1997. *Alliance Politics*. Ithaca: Cornell University Press.

Sorokin, Gerald L. 1994. "Arms, Alliances, and Security Tradeoffs in Enduring Rivalries." *International Studies Quarterly* 38(3):421–446.

Starr, Harvey. 1974. "A Collective Goods Analysis of the Warsaw Pact after Czechoslovakia." *International Organization* 28(3):521–532.

Starr, Harvey. 2000. "Substitutability in Foreign Policy: Theoretically Central, Empirically Elusive." *Journal of Conflict Resolution* 44(1):128–138.

Steenbergen, Marco R and Bradford S Jones. 2002. "Modeling Multilevel Data Structures." *American Journal of Political Science* pp. 218–237.

Thies, Wallace J. 1987. "Alliances and Collective Goods: A Reappraisal." *Journal of Conflict Resolution* 31(2):298–332.

Thompson, Alexander. 2010. "Rational design in motion: Uncertainty and flexibility in the global climate regime." *European Journal of International Relations* 16(2):269–296.

Thompson, William R. and David R. Dreyer. 2012. *Handbook of International Rivalries 1494–2010*. Washington, DC: CQ Press.

Thyne, Clayton L. 2012. "Information, Commitment, and Intra-War Bargaining: The Effect of Governmental Constraints on Civil War Duration." *International Studies Quarterly* 56(2):307–321.

Tomz, Michael. 2007. *Reputation and International Cooperation: Sovereign Debt Across Three Centuries*. Princeton: Princeton University Press.

Tomz, Michael, Judith L Goldstein and Douglas Rivers. 2007. "Do We Really Know That the WTO Increases Trade? Comment." *The American Economic Review* 97(5):2005–2018.

Trager, Robert F. 2010. "Diplomatic Calculus in Anarchy: How Communication Matters." *American Political Science Review* 104(02):347–368.

Treier, Shawn and Simon Jackman. 2008. "Democracy as a Latent Variable." *American Journal of Political Science* 52(1):201–217.

Vehtari, Aki, Andrew Gelman and Jonah Gabry. 2017. "Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC." *Statistics and Computing* 27(5):1413–1432.

Walt, Stephen M. 1990. *The Origins of Alliance*. Cornell University Press.

Walt, Stephen M. 2009. "Alliances in a Unipolar World." *World Politics* 61(01):86–120.

Warren, T Camber. 2016. "Modeling the coevolution of international and domestic institutions: Alliances, democracy, and the complex path to peace." *Journal of Peace Research* 53(3):424–441.

Weeks, Jessica L. 2008. "Autocratic audience costs: Regime type and signaling resolve." *International Organization* 62(1):35–64.

Weeks, Jessica LP. 2014. *Dictators at War and Peace*. Cornell University Press.

Whitten, Guy D. and Laron K. Williams. 2011. "Buttery Guns and Welfare Hawks: The Politics of Defense Spending in Advanced Industrial Democracies." *American Journal of Political Science* 55(1):117–134.

Williamson, Oliver E. 1985. *The Economic Institutions of Capitalism*. New York: Free Press.

Wolford, Scott and Moonhawk Kim. 2017. "Alliances and the High Politics of International Trade." *Political Science Research and Methods* 5(4):587–611.

Yarhi-Milo, Keren, Alexander Lanoszka and Zack Cooper. 2016. "To Arm or to Ally? The Patron's Dilemma and the Strategic Logic of Arms Transfers and Alliances." *International Security* 41(2):90–139.

Zielinski, Rosella Cappella, Benjamin O Fordham and Kaija E Schilde. 2017. "What goes up, must come down? The asymmetric effects of economic growth and international threat on military spending." *Journal of Peace Research* 54(6):791–805.

APPENDIX A

APPENDIX TO CHAPTER 3

This online appendix provides more detail about the multilevel model and checks the results. I also briefly describe a single-level test of the depth hypothesis and assess other measures of alliance treaty depth.

A.1 Descriptive Statistics of Key Variables

This section provides two tables of descriptive statistics for the state and alliance-level variables.

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
International War	8,280	0.029	0.167	0	0	0	1
Civil War Participant	8,280	0.082	0.275	0	0	0	1
Rival Military Spending	8,280	−0.062	0.432	−0.263	−0.263	0.101	3.444
GDP growth	8,280	0.008	0.516	−3.699	−0.187	0.184	20.202
POLITY	8,280	−0.011	0.502	−0.738	−0.532	0.496	0.633
Cold War	8,280	0.503	0.500	0	0	1	1
Annual MIDS	8,280	−0.060	0.384	−0.245	−0.245	0.144	9.869

Table A.1: State-level variables.

A.2 Priors

Table A.3 summarizes the prior distributions in the multilevel model. All priors are weakly informative relative to the scale of the data. ν is the degrees of freedom for the t-distribution, and the gamma prior is the recommended prior for STAN (Juárez and Steel, 2010).

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Latent Depth Mean	190	0.069	0.898	−0.843	−0.776	0.878	1.979
Unconditional Military Support	190	0.521	0.501	0	0	1	1
Economic Issue Linkages	190	0.579	0.495	0	0	1	1
Foreign Policy Concessions	190	0.805	0.860	0	0	1	3
Number of Members	190	3.879	5.677	2	2	3	43
Foreign Policy Disagreement	190	0.653	0.346	−0.103	0.360	0.972	1.000
Average Democracy	190	−2.127	5.708	−10	−7	1.5	10
Wartime Alliance	190	0.126	0.333	0	0	0	1
Asymmetric Obligations	190	0.216	0.412	0	0	0	1
Average Threat	190	0.356	0.163	0.000	0.253	0.473	0.674
US Membership	190	0.095	0.294	0	0	0	1
USSR Membership	190	0.074	0.262	0	0	0	1

Table A.2: Alliance-level variables. Number of members, foreign policy disagreement, average democracy, and average threat are all measured for the first year of the alliance.

$$
\begin{gathered}
p(\alpha) \sim N(0,1) \\
p(\sigma) \sim \text{half-}N(0,1) \\
p(\alpha^{yr}) \sim N(0,\sigma^{yr}) \\
p(\sigma^{yr}) \sim N(0,1) \\
p(\alpha^{st}) \sim N(0,\sigma^{st}) \\
p(\sigma^{st}) \sim \text{half-}N(0,.5) \\
p(\sigma^{all}) \sim \text{half-}N(0,.5) \\
p(\beta) \sim N(0,.5) \\
p(\gamma) \sim N(0,.5) \\
p(\nu) \sim Gamma(2,0.1)
\end{gathered}
$$

Table A.3: Summary of priors in the multilevel model

A.3 Hamiltonian Monte Carlo Diagnostics

There were no divergent iterations in either sample running 4 chains for 2,000 iterations with 1,000 warmup iterations. The $\hat{R}$ is less than 1.1 for all parameters in both samples. Trace plots in Figure A.1 indicate good mixing of the chains for the alliance-level parameters. Taken together, all of this implies that the chains adequately explored the posterior distribution.

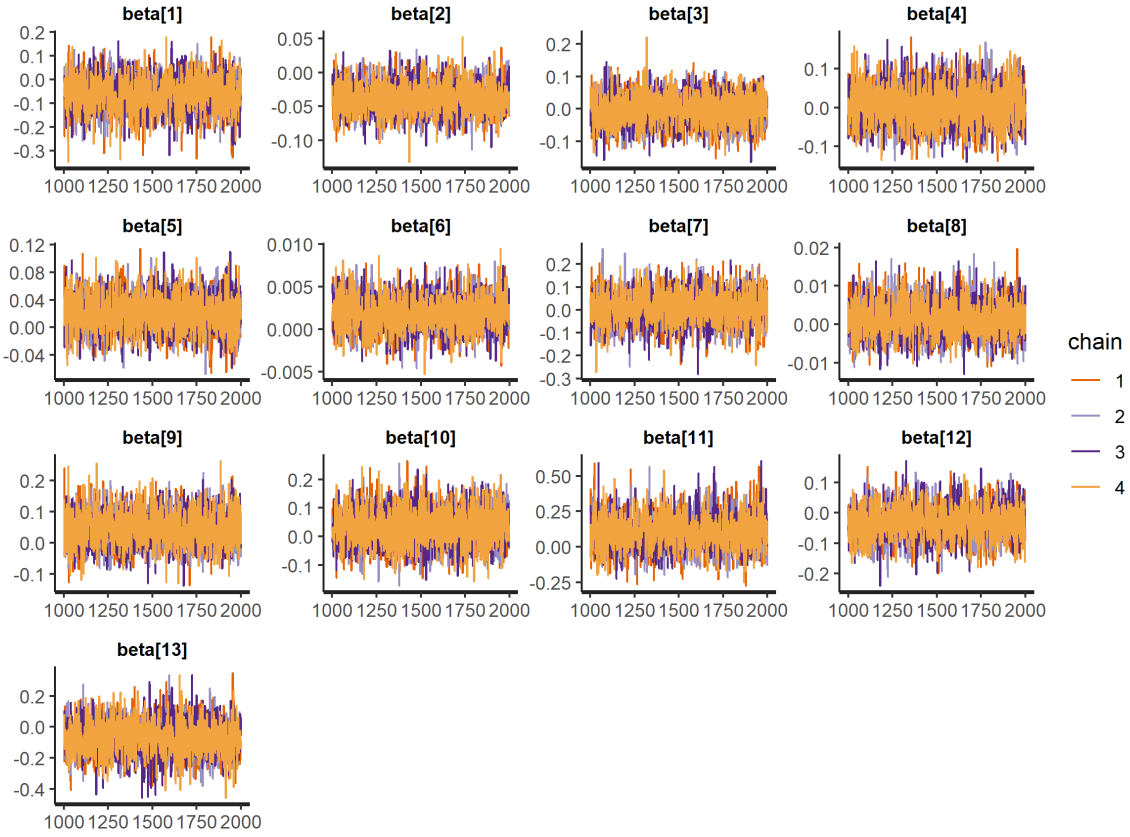

Figure A.1: Traceplot of alliance level parameters in the non-major power sample.

A.4 Normalizing Allied Capability

As noted in the paper, I place allied capability in the membership matrix **Z** on the same scale as the other parameters by normalizing it by year. Within each year, I divide the total military

spending of allied states by the maximum value, so capability values within each year range from just above zero to one. This ensures that allied capability is comparable within years, and that I do not treat more modern alliances as the most capable due to increases in raw defense budget sizes.

The choice of this specific normalization is less theoretically informed than using capability itself. Therefore, I assessed the use of different normalizations and rescalings for allied capability by comparing model fit. I fit three models in addition to the one presented in the paper. The first rescaled allied capability by dividing each capability value by the maximum of capability without grouping alliances by year. The second rescaled alliance capability by dividing by two standard deviations, which is problematic because it introduces negative capability values. The last used total allied CINC scores instead of military spending as an indicator of allied capability, which also facilitates comparisons of allied capability within years. CINC scores measure the share of total world military capability each state has in a particular year, so it is useful for comparing allied capability within years (Singer, 1988).

After estimating these three models, I used leave-one-out (LOO) cross validation to assess model fit (Vehtari, Gelman and Gabry, 2017). LOO estimates pointwise out-of-sample prediction accuracy using the log-likelihood evaluated at the posterior simulations of the parameter values.[1] All diagnostics indicate the LOO results are not driven by unusual observations. As with other information criteria, lower values indicate better fit.

Allied Capability	elpd_diff	se_diff	elpd_loo	se_elpd_loo
Normalized by Year	0.000	0.000	-1159.513	184.714
Rescaled by Maximum	-3.165	2.643	-1162.679	184.723
Recaled by 2SD	-10.749	6.116	-1170.262	184.741
Total Allied CINC	-12.308	5.576	-1171.821	184.683

Table A.4: Leave-one-out cross validation to assess model fit with different rescalings or normalizations of alliance capability.

[1]The widely applicable information criteria (WAIC) produces similar results, but the estimates for the CINC model may be driven by an unusual observation.

Table A.4 summarizes the assessment of each model using the expected log pointwise predictive density (elpd). I use the model from the paper as the comparison model: a negative elpd_diff implies the normalized model fits the data better. The difference also has some uncertainty, which is summarized by the se_diff column of Table A.4. The other three models have a negative elpd_diff compared to the model with normalized capability by year. For the models with CINC and rescaling by two standard deviations the difference is large, relative to the se_diff, so there is a clear preference for the normalized model. Normalizing by year provides at best a marginal improvement over a model where capability is rescaled using the maximum. Rescaling capability by the maximum produces similar inferences about alliance characteristics, including treaty depth.

A.5 Fake Data Simulation Check

With any complicated model, simulating fake data and seeing if the model can recover known parameters is essential. Fake-data simulation helps validate results from observed data and identify problems. This section summarizes results from fitting the multilevel model to fake data.

I simulated a dataset of 2000 t-distributed observations with 50 states observed for 200 years and 100 alliances. The outcome has a different scale than the military spending outcome variable, so coefficient values here do not match the paper. I then simulated two state and alliance level variables and took a piece of the matrix of state membership in alliances. Last, I ran the model without evaluating the likelihood, generating a posterior prediction of the outcome based on the fake data.

To check whether the model could recover known parameters, I took the 12th draw of the posterior distribution. This draw included a simulated outcome for each observation and a set of coefficients. I then fit the multilevel model on the simulated outcome values and checked whether the credible intervals contained the corresponding parameter values. If a parameter is within the 90% credible interval, the model captures it.

The model recovers known parameters with a high degree of accuracy. As shown by Figure A.2, The two credible intervals of the alliance-level regression include the known values.

Credible interval coverage for the variance hyperparameters and γ parameters is also acceptable.

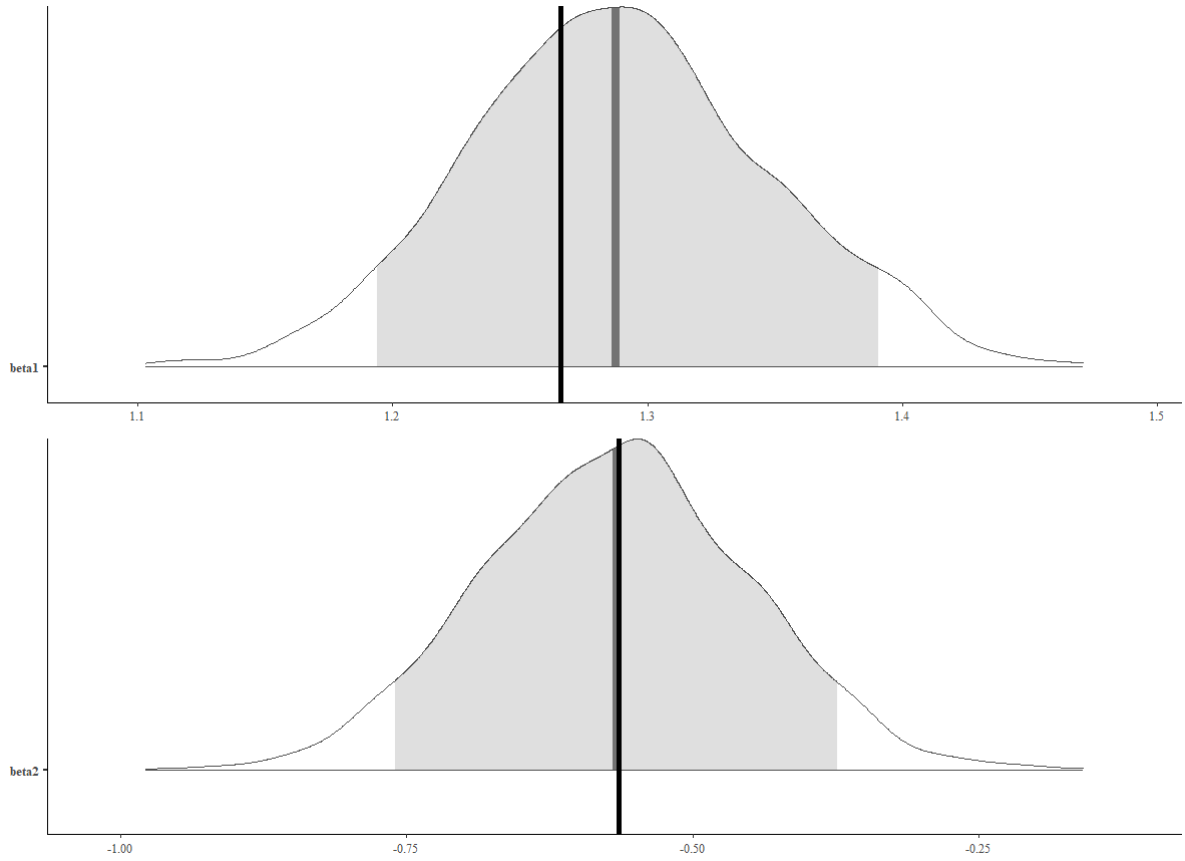

Figure A.2: Posterior distributions of β parameters from fitting multilevel model to fake data. The black vertical line marks the known parameter value, and the grey area is the 90% credible interval.

Even with small multiples, the 100 λ parameters are hard to plot, so I offer a descriptive summary here. Among the λ parameters, 93 of 100 intervals contain the known λ value. Given the large number of parameters and smaller sample, this is acceptable accuracy. Even the seven inaccurate confidence intervals were quite close— all were within .015 of the known parameter.[2]

[2]Fine margins around these intervals implies that the exact number of accurate λ intervals is sensitive to simulation variance.

A.6 Robustness Check: Alternative Measure of Military Spending

The main findings in the manuscript rely on the Correlates of War military spending. Due to reporting issues, definition problems and measurement challenges, other measures of military spending could lead to different results. I check the robustness of my results by using Nordhaus, Oneal and Russett (2012)'s measure of military spending, which combines data from the COW project and the Stockholm International Peace Research Institute (SIPRI). DiGiuseppe and Poast (2016) use this measure of military spending in their paper.

I estimate the same multilevel model on this measure of military spending, which covers from 1949 to 2001. This model also checks whether how treaty depth modifies the impact of alliance participation on military spending changes after World War II. Because the coefficient on a lagged dependent variable in this model is close to one, implying probable non-stationarity in levels, I use changes in military spending as the outcome of interest.

Figure A.3 summarizes the alliance-level regression parameters. As with the COW data, the credible interval for treaty depth is negative and does not overlap zero. All the parameter estimates are similar in this data, which increases my confidence that the results are not driven by the COW spending data.

A.7 Robustness Check: Single-Level Regression

Though the multilevel model best reflects the theory, I also fit some more standard panel data models. In what follows, I briefly present results from robust regressions of state-year percentage changes in military spending in the same sample of non-major powers. As in the multilevel model, I applied the inverse hyperbolic sine transformation to the outcome. In these models, I employ two indicators of alliance depth. The first is the average depth of a state's alliances. The second is a dummy which equals 1 if a state has at least one alliance with greater than average depth. Both variables compare states with different depth in their alliance portfolio. In addition to the state-level controls in the multilevel model, I included average alliance size, average allied democracy and the log of total allied capability as controls.

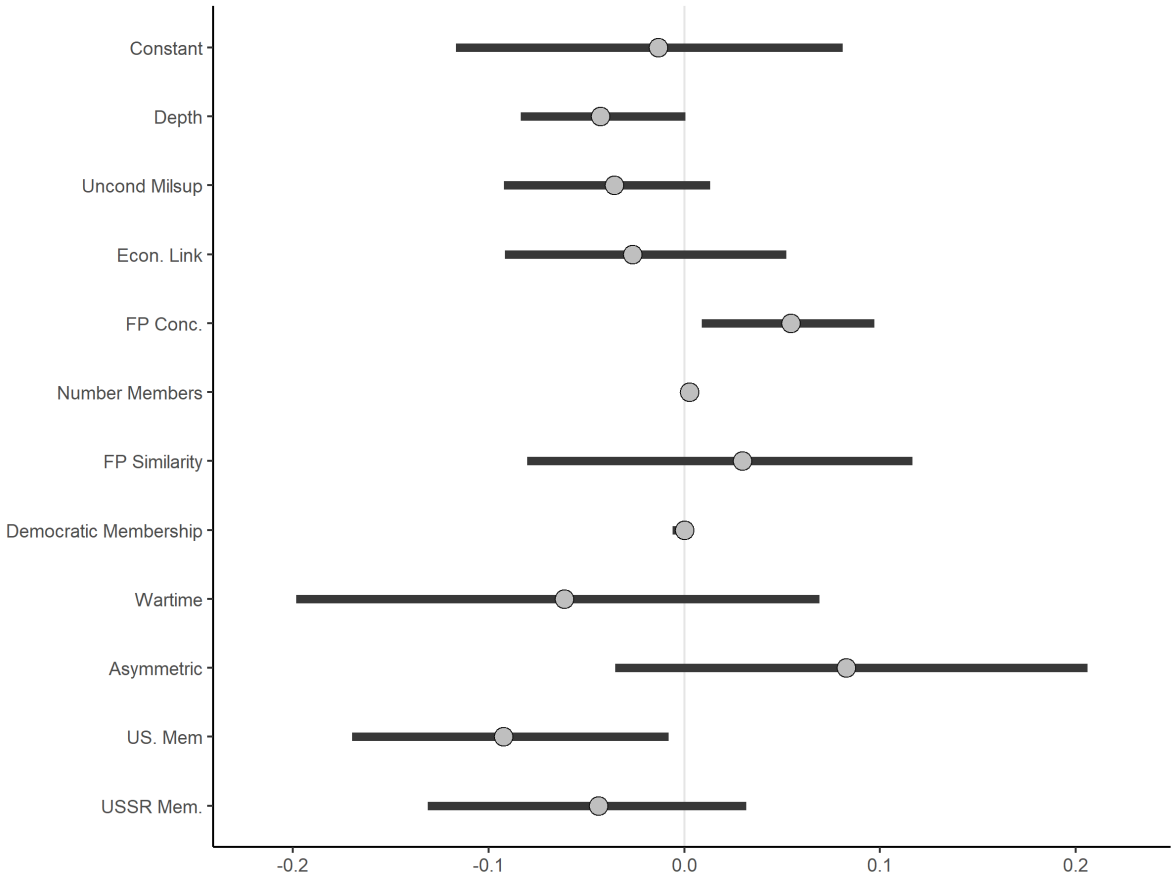

Figure A.3: 90% credible intervals of the β parameters from an analysis of changes in non-major power military spending from 1949 to 2001.

I estimated several models, including robust regressions on non-major powers and non-major powers in alliances. Comparing non-major powers with at least one alliance provides a crude approximation of the depth coefficient in the multilevel model, which compares deep and shallow alliances. I also applied state and year fixed effects to an OLS model of percentage changes in defense expenditures. The estimated association between average treaty depth and military spending changes is summarized in Figure A.4. Results are inconsistent- I do not reject the null hypothesis for the average depth measure unless the model includes state and year fixed effects. The deep alliance dummy coefficient estimate is negative and statistically significant across all samples and model specifications, however. The dummy variable is closest to DiGiuseppe and Poast (2016)'s design.

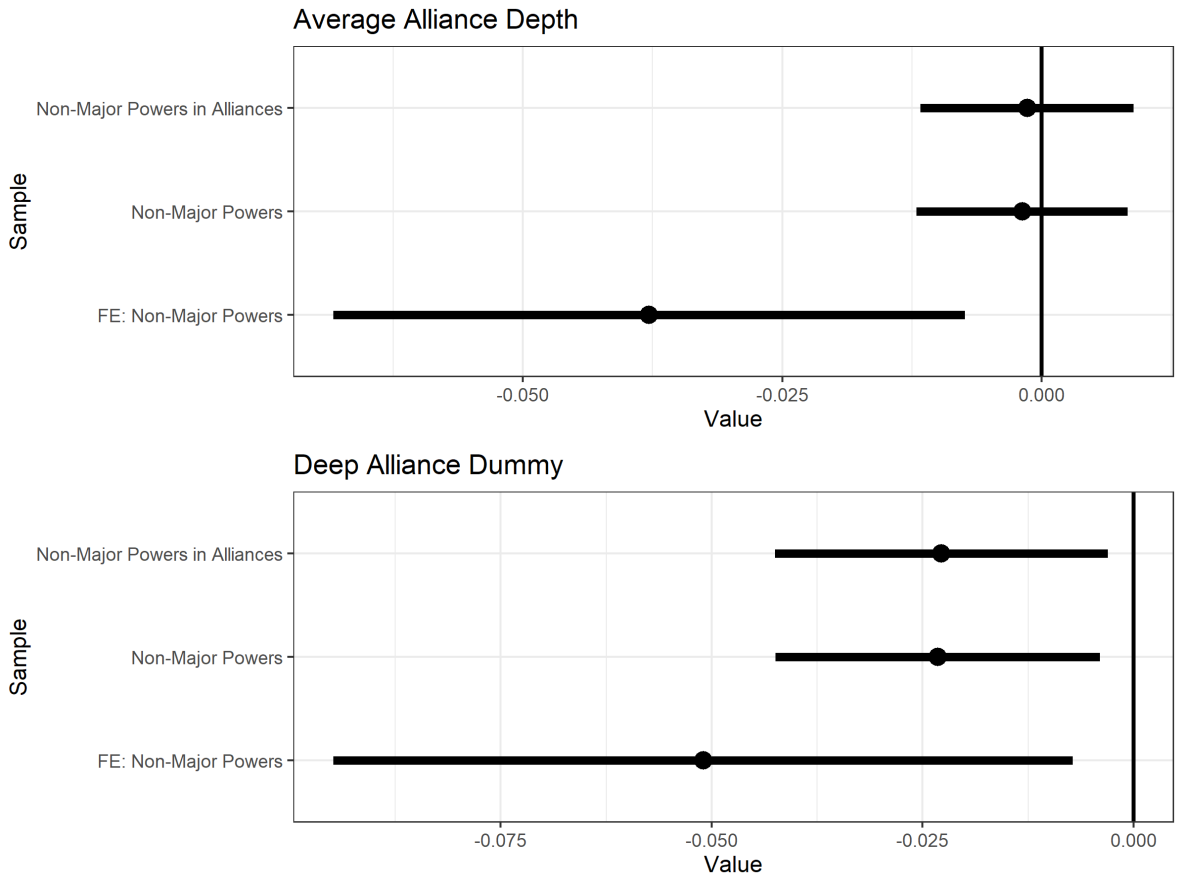

Figure A.4: Estimated effect of average alliance treaty depth or a dummy indicator of participation in a deep alliance on percentage changes in non-major power military spending from 1816 to 2007.

A.8 Alternative Measures of Alliance Treaty Depth

This part of the appendix compares my latent measure of treaty depth with similar measures in earlier research. I first show that there are important differences between my measure and the military institutionalization measure of Leeds and Anac (2005), but using military institutionalization instead of latent depth generates similar inferences about how treaty depth modifies the impact of alliance participation. Then I describe the conceptual and empirical differences between my latent measure and that of Benson and Clinton (2016).

A.8.1 Leeds and Anac 2005

Leeds and Anac (2005) create an ordinal measure of alliance treaty depth to study whether alliance institutionalization improves treaty reliability and performance in offensive and defense alliances. They argue that commitments of an integrated military command, common defense policy, or any basing rights generate high military institutionalization. Official contact between military officials, formal organizations, providing training or technology, subordination of forces, or specific contributions reflect moderate institutionalization. If at least one factor is present, Leeds and Anac assign alliance the highest corresponding level of institutionalization. This approach assumes that alliances with multiple sources of depth as just as deep/institutionalized as alliances with one factor and understates the amount of variation in alliance treaty depth.

Even so, as Figure A.5 shows, this ordinal measure and my latent measure are positively correlated. The deepest alliances on the latent measure also have the highest military institutionalization score, because they rely on similar variables. There are substantial differences within each category and overlap in the latent scores across the categories, however. For example, some alliances that Leeds and Anac (2005) assign a moderate institutionalization score have more depth than alliances with high institutionalization scores because these alliance treaties contain multiple sources of depth.

There are two sets of alliances where my measure makes a marked departure from Leeds and Anac. First, there are some alliances with no institutionalization that my measure assigns some

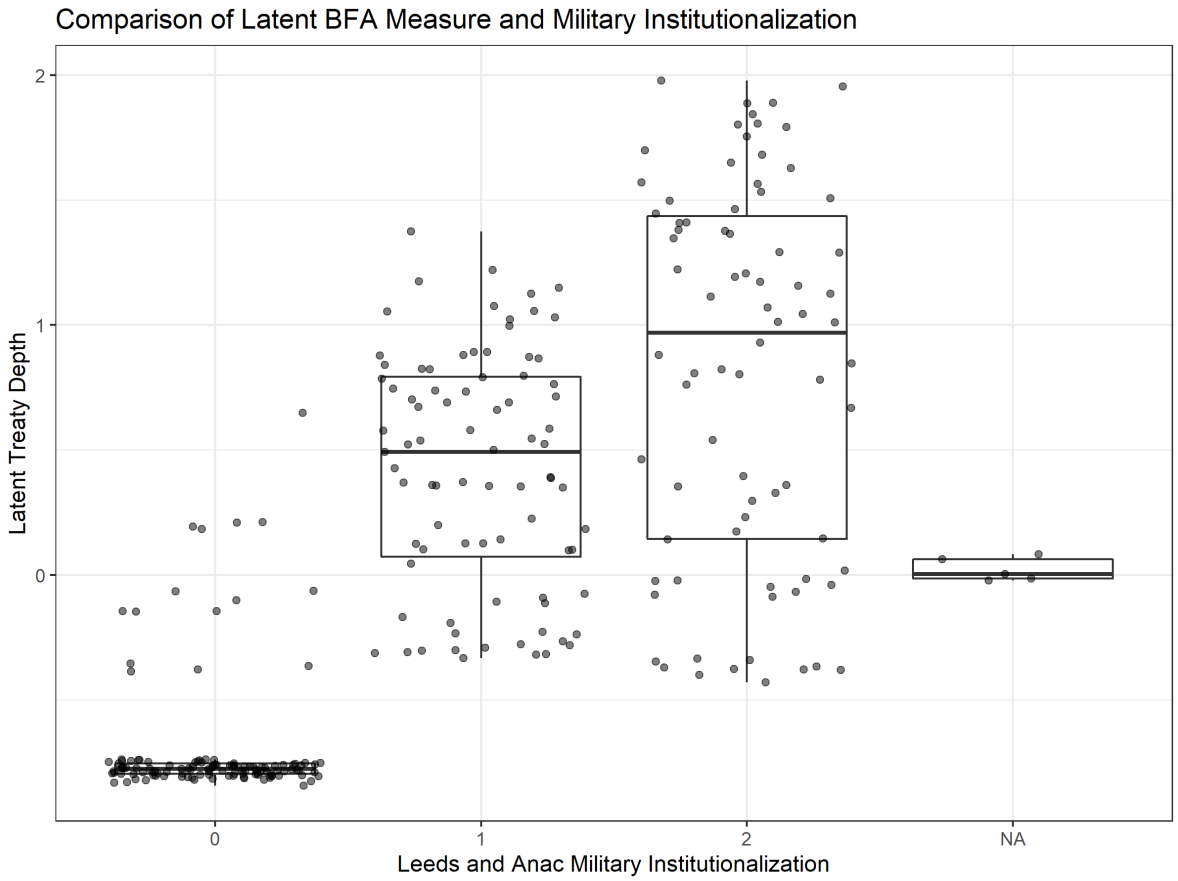

Figure A.5: Scatter plot of latent treaty depth across the values of military institutionalization from Leeds and Anac (2005). The box plots summarize the distribution of latent treaty depth within each category of military institutionalization. Points are jittered within each level of the institutionalization score.

depth to. This difference is the result of companion military agreements, which I include as a source of depth in addition to Leeds and Anac's variables. Second, Leeds and Anac assign missing values to some institutionalization scores if all sources of high or moderate depth are missing, which is a reasonable choice with their measurement strategy. My latent measure gives these treaties some depth, because it accommodates missing data on a subset of variables.

Although the latent depth measure has some advantages, I find similar results with the ordinal measure of military institutionalization. To check the robustness of my results, I implemented the same multilevel model of non-major power military spending, but replaced the mean latent treaty depth variable with the military institutionalization measure. Table A.5 summarizes the results.

	mean	sd	5%	95%	n_eff	$\hat{R}$
Constant	0.003	0.058	-0.094	0.096	2688.743	1.001
Military Inst.	-0.035	0.024	-0.075	0.004	3960.366	1.000
Uncond. Milsup.	-0.018	0.042	-0.087	0.051	3450.201	1.000
Econ. Link	0.015	0.049	-0.066	0.096	3056.525	1.000
FP Conc.	0.025	0.025	-0.017	0.066	4115.104	1.000
Number Members	0.002	0.002	-0.001	0.004	3696.671	1.000
FP Similarity	-0.005	0.065	-0.110	0.105	2697.860	1.001
Democratic Membership	0.001	0.004	-0.006	0.008	3134.146	1.000
Wartime	0.047	0.051	-0.037	0.132	3879.252	1.000
Asymmetric	0.056	0.059	-0.037	0.155	2909.115	0.999
US. Mem	-0.033	0.049	-0.112	0.049	2617.047	1.000
USSR Mem.	-0.079	0.098	-0.237	0.083	3185.998	1.000
σ Alliances	0.143	0.054	0.060	0.234	914.843	1.003

Table A.5: Results from an analysis that replaces the latent measure of treaty depth with an ordinal measure of military institutionalization from Leeds and Anac (2005). The negative correlation between military institutionalization and the impact of alliance participation on military spending matches earlier conclusions about the way treaty depth impacts military spending.

The same finding about treaty depth holds when the analysis uses military institutionalization in place of mean latent treaty depth. Military institutionalization and the impact of alliance participation on military spending are negatively correlated. 93% of the posterior probability in the depth coefficient is negative, which matches Hypothesis 3.

Figure A.6 helps assess Hypotheses 1 and 2. Among alliances with no institutionalization, most treaties have a positive effect. Alliances with moderate institutionalization have mixed effects. Last, participation in alliances with high institutionalization tends to reduce military spending. This corresponds to the predictions of Hypotheses 1 and 2. The trend across military institutionalization is less clear than with the latent measure of treaty depth, probably because the military institutionalization measure understates variation in treaty depth.

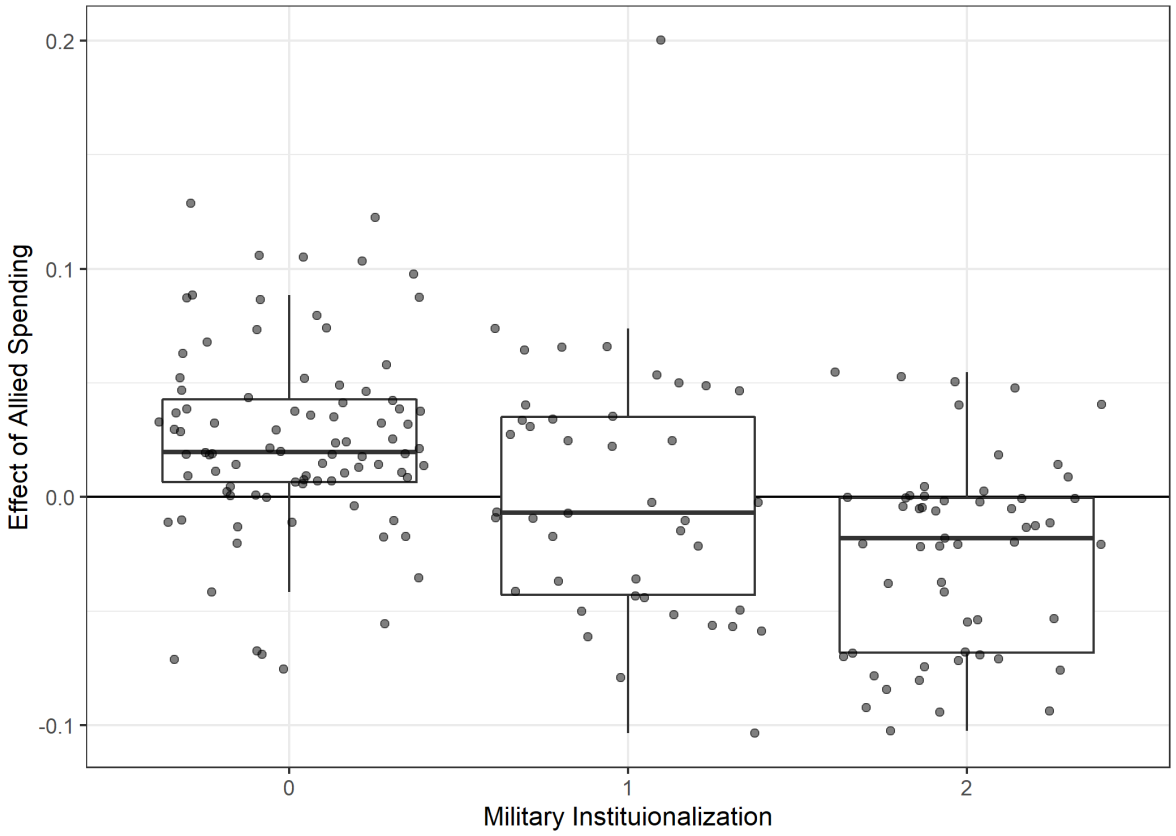

Figure A.6: Scatter plot of λ parameters against the values of military institutionalization. The box plots summarize the distribution of points within in each military institutionalization value. All points are jittered to make the plot more legible.

A.8.2 Benson and Clinton 2016

Having established that an ordinal measure of treaty depth produces similar results, I now explain why I did not use an existing latent measure of treaty depth. Benson and Clinton (2016) use a latent variable model to estimate alliance scope, depth and capability. I do not use their measure of depth because it captures a different concept, and thus diverges from my aims in this project. Benson and Clinton define depth as the general costliness of alliance obligations, which leads them to include other variables and measure the depth of neutrality pacts. I define depth in terms of military coordination and cooperation and am only interested in alliances with active military support. My focus on defensive and offensive alliances follows existing scholarship on alliance participation and military spending. These conceptual differences, along with my use of a different estimator, lead to different conclusions about the factor loadings and the latent depth scores.

Benson and Clinton aim for a broad measure of alliances, so they include neutrality pacts in their data. There is an understandable choice, but it means their measure diverges from the my argument, which focuses on alliances with military support. Neutrality pacts are qualitatively different, because peacetime coordination is not focused on ensuring the delivery of military support.

In Benson and Clinton's measure, neutrality pacts have very little depth, which is unsurprising. Neutrality is less costly in general. Only a few alliances with only neutrality obligations have any depth, as Figure A.7 shows.

In estimation, neutrality pacts are something like a reference category for military alliances. Neutrality pacts have limited depth, but they inform inferences about the depth of alliances with offensive or defensive promises. By including neutrality pacts, Benson and Clinton's model compares alliances with generally limited obligations to alliances with active military support.

Again, Benson and Clinton's decision to measure the depth of neutrality pacts is not a problem for their paper. It limits the applicability of their measure to this project, however. I do not address neutrality pacts in my argument, so my measure of depth only considers variation among offensive and defensive alliances. My theoretical focus on military cooperation also leads me to exclude in-

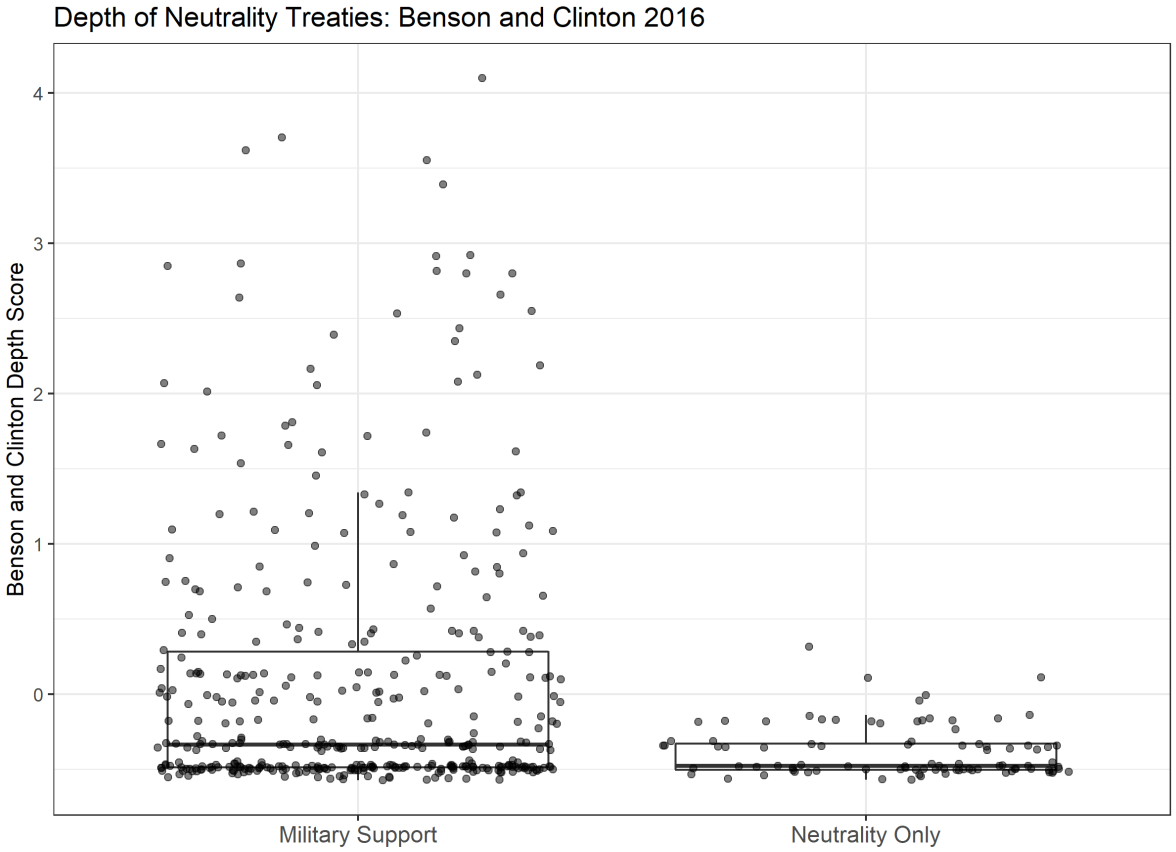

Figure A.7: Comparison of Benson and Clinton's depth measure among alliances with only neutrality obligations and alliances with active offensive or defensive support. The box plots summarize the distribution of alliance depth within each group of alliances. All points are jittered to make the plot more legible.

dicators of secrecy and economic aid that Benson and Clinton use to predict treaty depth. Although Benson and Clinton's measure is close to my purposes, it captures a slightly different concept.

Given conceptuial differences and my decision to use a semiparametric estimator that breaks problematic correlations between the latent variables and dependence structure (Murray et al., 2013), I draw different conclusions about the factor loadings and distribution of treaty depth. Figure A.8 describes the key differences between my latent measure of depth and Benson and Clinton's measure. In the top panel, I look at differences in the factor loadings across the two models.[3] Benson and Clinton break the ATOP policy coordination variable into military contact and common defense policy, but I treat this as an ordinal variable, which is the largest source of depth.[4] I also find larger correlations between formal organization and integrated military command and latent depth than Benson and Clinton. It is harder to distinguish between the other loadings, but Benson and Clinton's measure assigns marginally more weight to military aid and basing rights.

In the bottom panel of Figure A.8, I plot my measure of treaty depth against Benson and Clinton's. To facilitate this comparison, I rescaled both depth measures by dividing them by one standard deviation. Benson and Clinton's measure suggests 18 alliances are 2 or more standard deviations from the mean, while my measure contains 9 such alliances. Many treaties with high depth on Benson and Clinton's measure have lower depth relative to other alliances on my measure. The two measures identify a common set of six extremely deep alliances, but disagree about how distinctive they are from other treaties. Salient distinctions in the relative depth of alliance treaties follow from differences in the factor loadings.

Though the two measures of depth are correlated, they capture different concepts and my measure has fewer extreme outliers. Benson and Clinton address the general cost of an alliance, while I am focused on military coordination and cooperation. Benson and Clinton's measure is useful, but it departs from my aims in this project by including other variables and analyzing neutrality pacts.

[3]Recall that I removed economic aid and secrecy from the observed data in my measure, because they are distinct from military cooperation.

[4]The military contact variable on which the policy coordination score is based gives alliances that require wartime military contact a score of one, scores treaties with peacetime military contact as a two, and assigns alliances with defense cooperation a score of three. There is a order to the extent of policy coordination required as this variable increases.

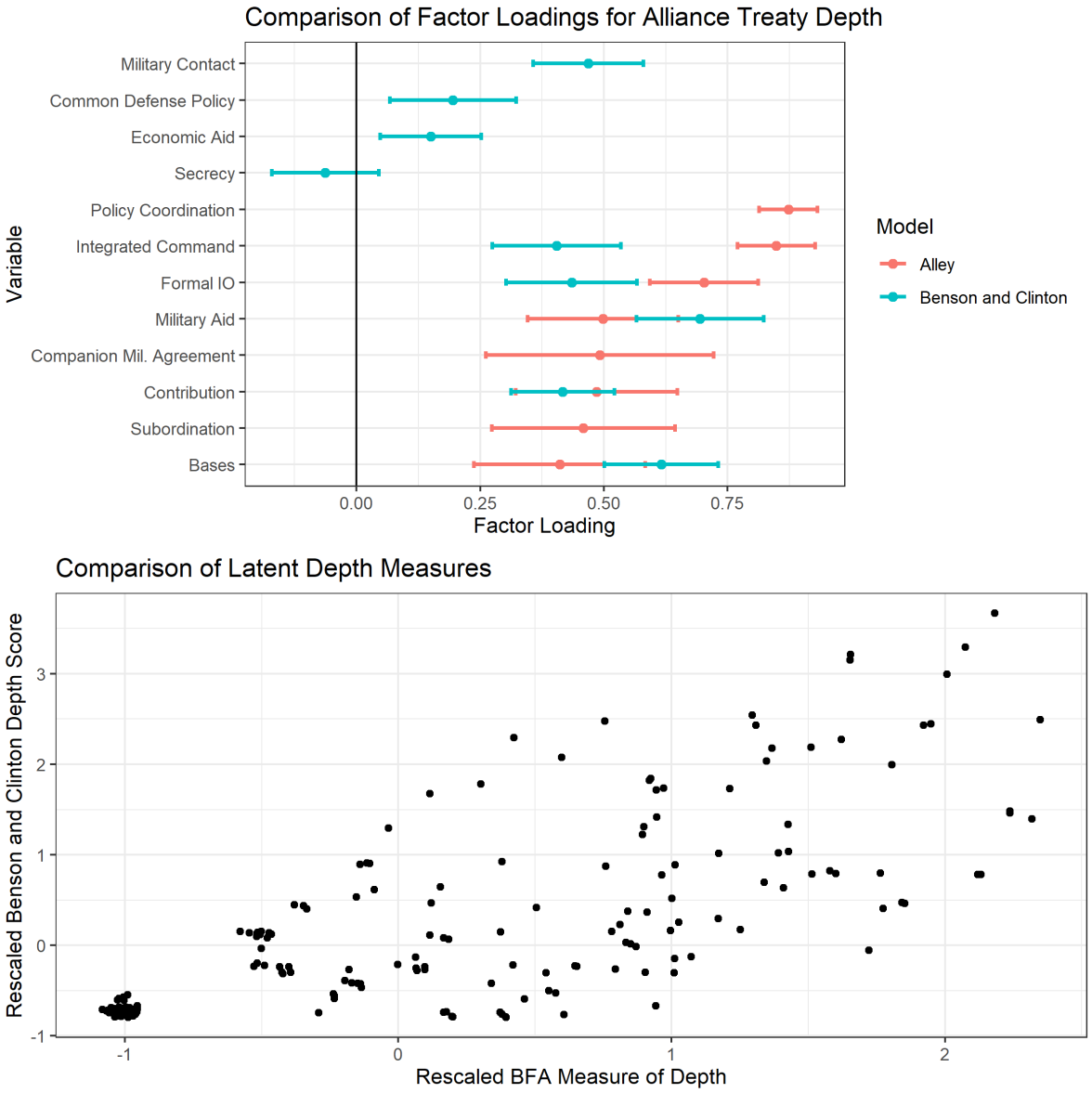

Figure A.8: Comparison of latent measures of alliance treaty depth, one from this paper, and the other from Benson and Clinton (2016). The top panel compares the factor loadings from the two variables. The scatter plot compares latent depth scores from the semiparametric factor analysis with depth scores from Benson and Clinton (2016). The two latent measures have been rescaled by one standard deviation to facilitate comparisons. This comparison only includes alliances from version 3 of the ATOP data, because these are the alliances for which both models have scores.

Because our measures operationalize different concepts in different groups of alliances, I believe my measure of depth is better suited for an analysis of the consequences of deep military cooperation in defensive and offensive alliances. My measure is not generally superior, and which latent depth measure scholars use in other analyses should depend on how they conceptualize alliance treaty depth.

APPENDIX B

APPENDIX TO CHAPTER 1

This appendix contains supporting materials for the test of Hypotheses 1 and 2 in "Reassessing the Public Goods Theory of Alliances." Section 1 assesses model fit and accuracy. The other section summarizes a test that uses average economic weight in their alliances to predict percentage changes in military spending.

B.1 Multilevel Model

This section describes the priors on the multilevel model, convergence diagnostics for the Hamiltonian Monte Carlo, and results from running the same model with a weighted economic size in the alliance participation matrix.

B.1.1 Priors

All priors are specified to be weakly informative relative to the scale of the data (Gelman, Simpson and Betancourt, 2017). I summarize the prior distributions for each set of parameters in Table B.1. $p(\nu)$ is a well-behaved prior for the degrees of freedom in a t-distribution (Juárez and Steel, 2010). Given that the median percentage change in military expenditures is 0.06, the priors are quite diffuse.

To facilitate estimation, I use a non-centered parameterization for the state and year varying intercepts, as well as the γ parameters (Betancourt and Girolami, 2015). A non-centered parameterization decouples the mean and variance to express an equivalent prior, which makes sampling easier. I also employ a sparse matrix representation of the alliance membership matrix $\mathbf{Z}$ to speed up estimation.

B.1.2 Convergence

There were no divergent iterations in sampling. However, there are other threats to inference from the posterior samples. Given heavy tails in percentage changes of military spending, STAN

$$
\begin{gathered}
p(\alpha) \sim N(0,1) \\
p(\sigma) \sim \text{half-}N(0,1) \\
p(\alpha^{yr}) \sim N(0,\sigma^{yr}) \\
p(\sigma^{yr}) \sim N(0,1) \\
p(\alpha^{st}) \sim N(0,\sigma^{st}) \\
p(\sigma^{st}) \sim \text{half-}N(0,1) \\
p(\gamma) \sim N(\theta,\sigma^{all}) \\
p(\theta) \sim N(0,.5) \\
p(\sigma^{all}) \sim \text{half-}N(0,1) \\
p(\beta) \sim N(0,1) \\
p(\nu) \sim Gamma(2,0.1)
\end{gathered}
$$

Table B.1: Summary of priors in multilevel model of Chapter 1.

might have struggled to explore the posterior distribution.

Energy plots can diagnose this problem. Figure B.1 plots the marginal energy distribution and the first differenced distribution. If the two histograms do not overlap, sampling was impeded by heavy tails. The substantial overlap in the distributions for all four chains in Figure B.1 indicates this was not a problem.

The split $\hat{R}$ statistic is another way to assess convergence. $\hat{R}$ compares the behavior of each chain by measuring the ratio of the average variance of draws within each chain to the variance of the pooled draws across chains. When $\hat{R}$ is close to 1, all the chains have similar variance, and are therefore in equilibrium.

The standard heuristic is that an $\hat{R}$ greater than 1.1 is problematic. Figure B.2 plots the $\hat{R}$ statistic for every parameter in the model. No parameters generate concern, even at a more conservative threshold of 1.05.

B.1.2.1 Inferences from Simulated Data

To assess if the model gives reasonable answers, I simulated data and associated parameters, then re-estimated the model on the simulated data. The model is a good fit if the credible intervals contain the known parameter values for the simulated data. This process checks whether the model can recover parameters from a known data-generating process that matches the model.

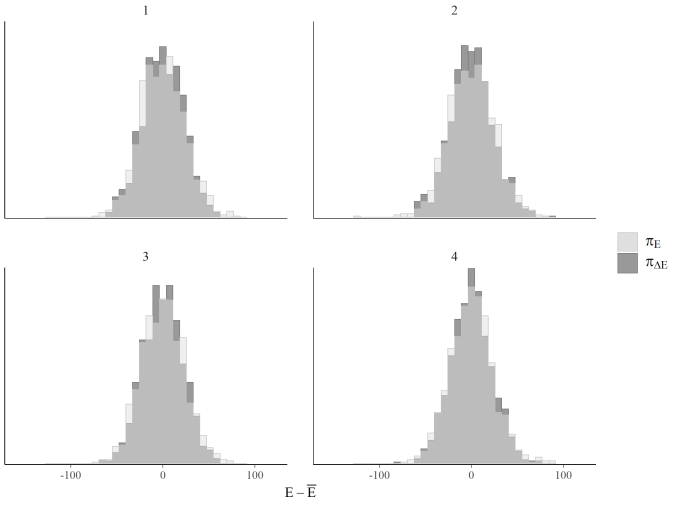

Figure B.1: Energy plot of multilevel model results. Greater overlap in the two histograms indicates adequate exploration of the posterior distribution.

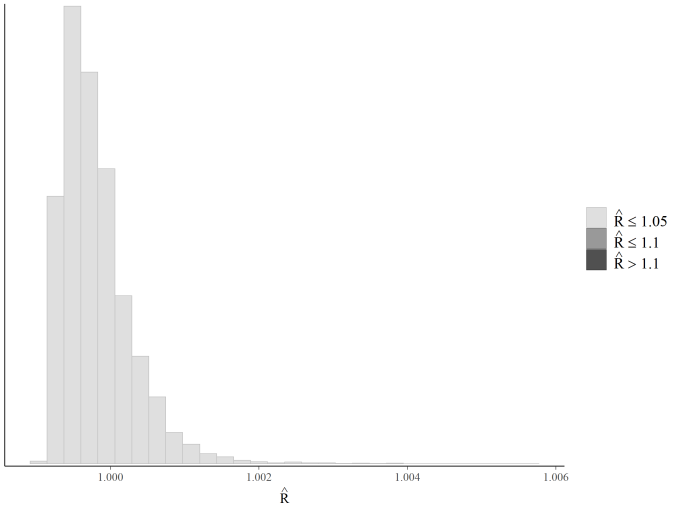

Figure B.2: Histogram of split $\hat{R}$ statistic for all parameters in the multilevel model.

I simulate a hypothetical dataset with 2000 observations of 50 states observed over 200 years. I used part of the observed alliance data from the paper, due to problems simulating an alliance membership matrix with the same characteristics as the alliance data. There are 100 alliances in this data along with 2 state-level control variables. The hypothetical outcome is drawn from a Cauchy distribution with mean 0 and a scale of .25, which is more heavy-tailed than even my observed data.

I then simulate 2,000 draws of the outcome using the generated quantities block in STAN. The next step is selecting one of those draws of the outcome— which includes the value of the outcome for each observation and the associated parameter values. I select the 12th draw from the posterior and check whether after estimating the model on these data, the credible intervals include zero.

I focus on inferences about the γ, θ and σ_{all} parameters, because all three affect my test of the public goods argument. As Figure B.3 and Figure B.4 show, the posteriors accurately capture the known values of the hyper-parameters θ and σ_{all}. In these figures, the true parameter value is marked with a thick black line, while the light gray shaded area shows the 90% credible interval.

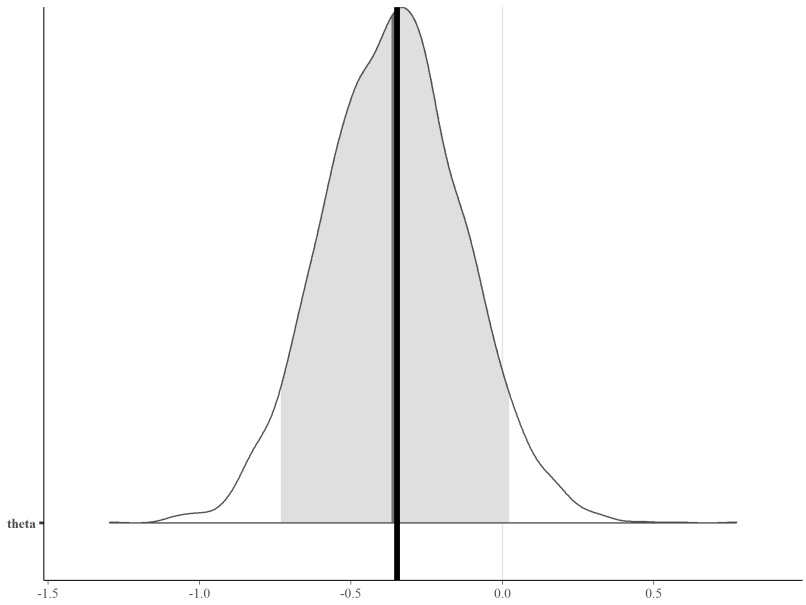

Figure B.3: Posterior estimates and known parameter value for the alliance hyperparameter θ. The dark gray bar marks the posterior mean, while the shaded area captures the 90% credible interval. The black line marks the known, "true" θ value, which falls within the 90% interval.

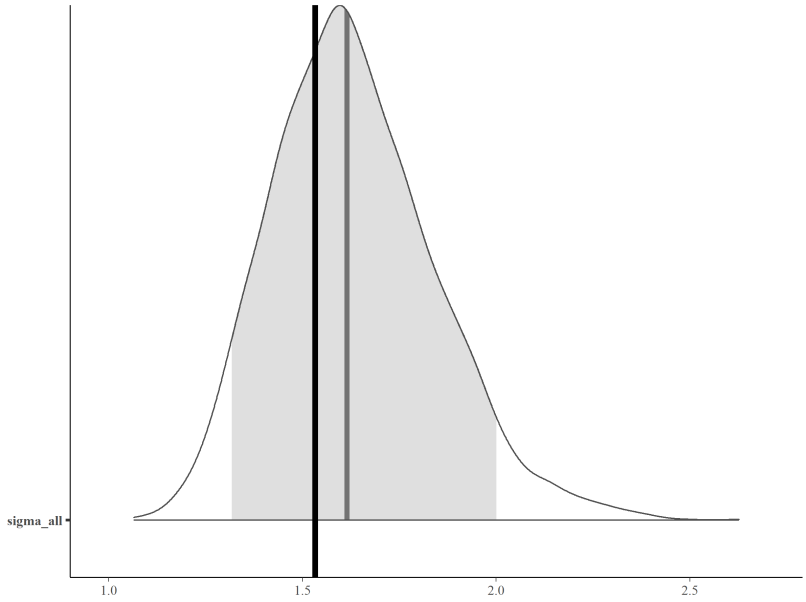

Figure B.4: Posterior estimates and known parameter value for the alliance hyperparameter σ_{all}. The dark gray bar marks the posterior mean, while the shaded area captures the 90% credible interval. The black line marks the known, "true" σ_{all} value, which falls within the 90% interval.

Because graphical presentation of the 100 γ parameters is more difficult, I calculated whether the credible interval contained the known parameter. 88 of the 100 intervals include the "true" γ value. Given the number of parameters and potential simulation variance, such accuracy is tolerable. Simulating data and recovering known parameters shows that the model estimates are reasonable approximations of the data-generating process.

B.1.3 Alternative Coding of Economic Size

The values of -1 for small states and 1 for large states in alliance participation matrix **Z** in the manuscript results create coarse bins. This could mask differences within each category that affect inferences about economic weight and percentage changes in military spending within alliances. This section checks whether inferences are sensitive to an alternative coding of **Z**.

To retain the same split of negative values for small states and positive values for large states, I subtracted one from economic weights that fell below the median in bilateral or multilateral alliances. This means that states with a small share of allied GDP have larger negative values than

states with close to the median value. For example, a state with a 25% of total allied GDP in a bilateral alliance has a weight of -.75 with this variable, but a state with 49% of allied GDP has a weight of -.51. Positive values like .51 are unchanged.

Again, the public goods model would predict many positive γ parameters, as these would reflect higher military spending for large states and lower military spending for small states. As Figure B.5 shows, there are few positive γ values. There are two alliances with a clearly positive γ parameter. These are the OAS (ATOPID 3150) and an alliance between the United States and Thailand (ATOPID 3260). 35 alliances have a positive posterior mean, as well.

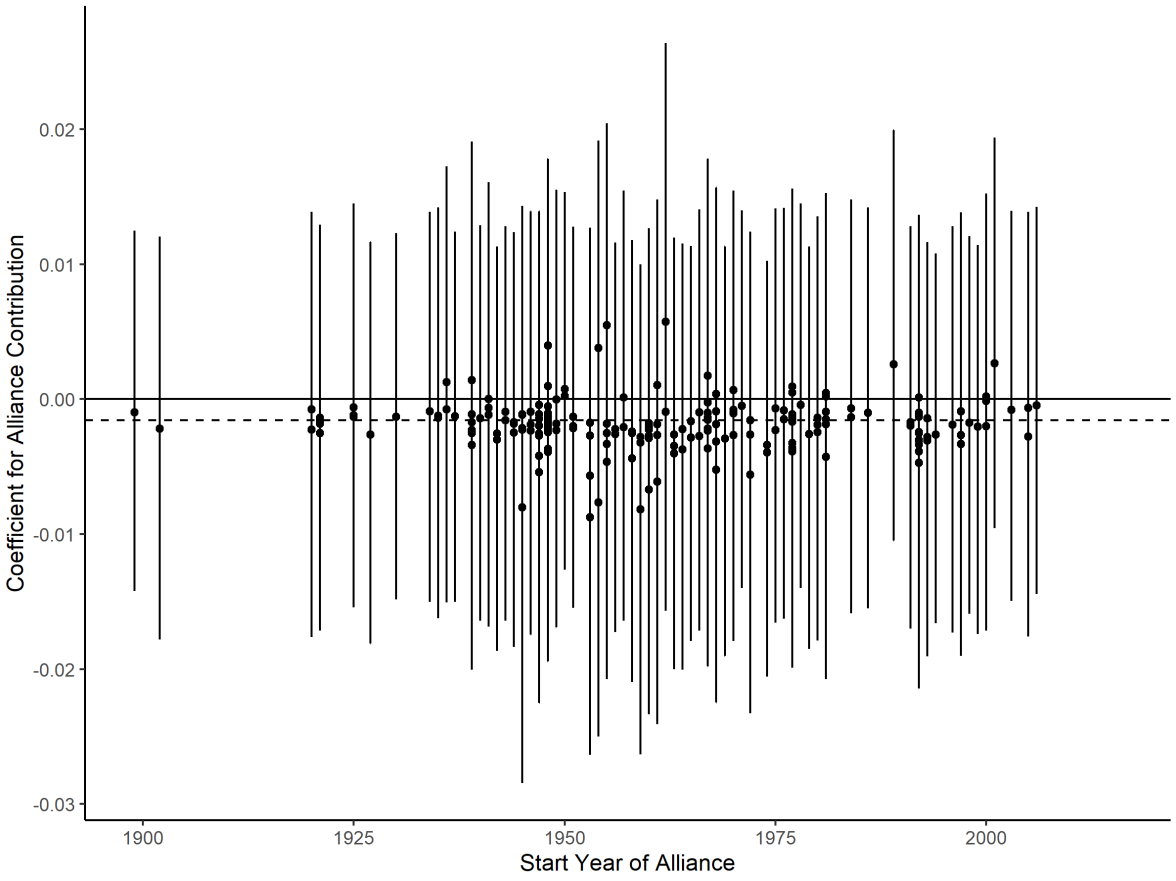

Figure B.5: Estimated γ parameters from a model with negative and positive economic weights in the alliance membership matrix.

While this approach is somewhat better for the public goods model's predictions, the evidence

is similar. There are very few alliances with even limited evidence that economic weight leads to higher military spending.

B.2 Average Economic Weight and Military Spending

The model I use in the paper estimates alliance-specific associations between economic weight and military spending. I also consider whether higher average economic weight across multiple alliances increases percentage changes in military spending. This section summarizes results from four models of the correlation between average economic weight and military spending. Following the recommendations of Rainey and Baissa (2020), I use OLS and robust regression estimators with both transformed and unaltered military spending growth. I use an inverse hyperbolic sine to transform the outcome because this transformation includes positive, negative and zero values. Robust regression down weights unusual observations, which is important because the residuals of OLS estimators in this data are not normally distributed.

The key independent variable is a state's average economic weight across all of its alliances. I also include measures of average alliance size and democracy (DiGiuseppe and Poast, 2016) as controls. Other controls are the same as the multilevel model.

Table B.2 summarizes the results of these four estimation strategies. The robust regression coefficients for average economic weight are of small substantive magnitude and the 95% confidence intervals include a range of positive and negative values. The OLS estimates for average economic weight are larger, but the difference in residual standard error between these estimates and the robust regression suggests that the OLS estimates are driven by unusual observations.

To asses the substantive impact of increasing a state's average economic weight in its alliances, I simulated the effect of moving average weight from the first quartile (0.05) to the third quartile (.40). Holding all other variables at their medians or modes, this increase in weight implies an expected percentage change in military spending of -0.005. The 90% credible interval for this change ranges from -0.017 to 0.01, so the estimated impact of a massive rise in average economic weight includes large positive, large negative and null effects.

	Dependent variable:			
	% Change Milex.		IHS(% Change Milex.)	
	robust linear	*OLS*	*OLS*	*robust linear*
	(1)	(2)	(3)	(4)
Avg. Economic Weight	−0.008	0.382	0.030	−0.008
	(−0.051, 0.036)	(−1.024, 1.789)	(−0.058, 0.118)	(−0.051, 0.036)
ln(GDP)	−0.035	−6.656	−0.363	−0.033
	(−0.135, 0.066)	(−9.896, −3.417)	(−0.566, −0.160)	(−0.133, 0.066)
Avg. Alliance Size	−0.0001	−0.014	0.00004	−0.0001
	(−0.001, 0.001)	(−0.039, 0.011)	(−0.002, 0.002)	(−0.001, 0.001)
Avg. Allied Democracy	0.0003	−0.031	−0.001	0.0003
	(−0.001, 0.002)	(−0.084, 0.022)	(−0.005, 0.002)	(−0.001, 0.002)
International War	0.082	0.149	0.140	0.080
	(0.054, 0.109)	(−0.728, 1.026)	(0.085, 0.195)	(0.053, 0.107)
Civil War Participant	−0.003	0.178	0.020	−0.003
	(−0.023, 0.017)	(−0.464, 0.821)	(−0.020, 0.061)	(−0.023, 0.017)
Regime Type	−0.001	0.039	0.0002	−0.001
	(−0.002, 0.001)	(−0.002, 0.079)	(−0.002, 0.003)	(−0.002, 0.001)
External Threat	0.072	1.109	0.107	0.071
	(0.040, 0.104)	(0.071, 2.146)	(0.042, 0.172)	(0.039, 0.103)
Cold War	0.040	0.439	0.040	0.040
	(0.029, 0.052)	(0.058, 0.820)	(0.016, 0.064)	(0.028, 0.052)
Constant	0.067	5.420	0.353	0.065
	(−0.015, 0.148)	(2.791, 8.049)	(0.189, 0.518)	(−0.016, 0.146)
Observations	5,022	5,022	5,022	5,022
R^2		0.007	0.015	
Adjusted R^2		0.005	0.013	
Residual Std. Error (df = 5012)	0.160	6.230	0.390	0.159
F Statistic (df = 9; 5012)		3.657 (p = 0.0002)	8.219 (p = 0.000)	
Note:			95% Confidence Intervals in Parentheses.	

Table B.2: OLS and robust regression of the association between average economic weight in alliances and percentage changes in military expenditures from 1919 to 2007.